Graduate School for Cellular and Biomedical Sciences
University of Bern

Genetic analysis of coat color phenotypes in domestic animals

PhD Thesis submitted by

Jan Wolfgang Henkel

from **Germany**

for the degree of

PhD in Computational Biology

Supervisor
Prof. Dr. Tosso Leeb
Institute of Genetics
Vetsuisse Faculty of the University of Bern

Co-advisor
Prof. Dr. Christine Flury
Department of Agronomy
School of Agricultural, Forest and Food Sciences HAFL

Accepted by the Faculty of Medicine, the Faculty of Science and the Vetsuisse Faculty of the University of Bern at the request of the Graduate School for Cellular and Biomedical Sciences

Bern, Dean of the Faculty of Medicine

Bern, Dean of the Faculty of Science

Bern, Dean of the Vetsuisse Faculty Bern

Table of Contents

1. Abstract

In many mammalian species, depigmentation phenotypes are widely known, but are rare in the wild, as unpigmented animals usually have a disadvantage in survival. Depigmentation phenotypes are sometimes associated with pleiotropic effects, resulting in severe pathological conditions such as e.g. the overo lethal white foal syndrome in horses. As abnormal coat color phenotypes are easily recognized in domestic animals they provide a good system to identify genetic variants affecting fundamental processes of life. This includes melanocyte development, migration, survival, and proliferation. Said special coat color phenotypes are mostly due to spontaneous mutations. The resulting mutant alleles, are often kept in the population and positively selected due to the striking appearance and special value of fully or partially depigmented animals to their owners.

In this thesis, I contributed to the identification of candidate causative genetic variants of various coat color phenotypes in a range of domestic animals, by applying different genetic approaches and with the help of next generation sequencing technologies. One approach was based on the calculation of selection signatures in various goat breeds from Switzerland and Pakistan. In two Pakistani goat breeds, complex structural variants at the *KIT* locus were identified potentially explaining the observed breed-defining depigmentation phenotypes. In the same study, multiple structural variants at the *ASIP* locus in Swiss goat breeds were found. RNA sequencing results of the caprine *ASIP* locus, revealed a high number of non-coding 5'-exons which were not observed in other mammals so far. Further analysis of one particular Swiss goat breed, the Valais Blackneck, revealed introgression events of some aforesaid *ASIP* alleles into the breed accounting for the rare coat color phenotypes segregating in the breed. Variation of *ASIP* could also be found in rabbits being a plausible cause for the black and tan phenotype further giving insight into the variable regulation of *ASIP* in multiple species. A heterozygous deletion affecting *KIT* was identified in a young German Riding Pony showing an interesting white spotting phenotype. It has most likely arisen by a *de novo* mutation event as the parents did not show any comparable phenotype. Lastly, a deletion in the *MITF* gene was found in a family of American Paint Horses, displaying variable white spotting phenotypes associated with deafness. Additional known white spotting alleles further increased the risk for deafness as six of the eight family members were deaf and carried multiple additional white spotting alleles.

The identification of these candidate causative variants not only enabled genetic testing and targeted breeding. They also gave insights into underlying mechanisms of the spatial and temporal regulation of pigment production as well as on the complex regulatory network

involved in melanocyte development. Additionally, they help us in better understanding of breed history.

2. Introduction

Wild animals have distinct coat color phenotypes, which are important for camouflage, hunting, mimicry, social communication, mating, predator warning and protection from UV light [1,2]. Depigmentation phenotypes are widely known in many mammalian species, but they are rare in wild animals, as unpigmented animals have a disadvantage in survival. Said special coat color phenotypes, including depigmentation, are mostly due to spontaneous mutations. In domestic animals the mutant alleles are often positively selected and kept in the population due to the special value of animals with striking appearance to their owners [3] (Figure 1). Depigmentation phenotypes result from the lack of melanocytes in hair and skin follicles [4] and the mutants provide a valuable resource for the genetic dissection of pathways that influence melanocyte development, migration, survival, and proliferation.

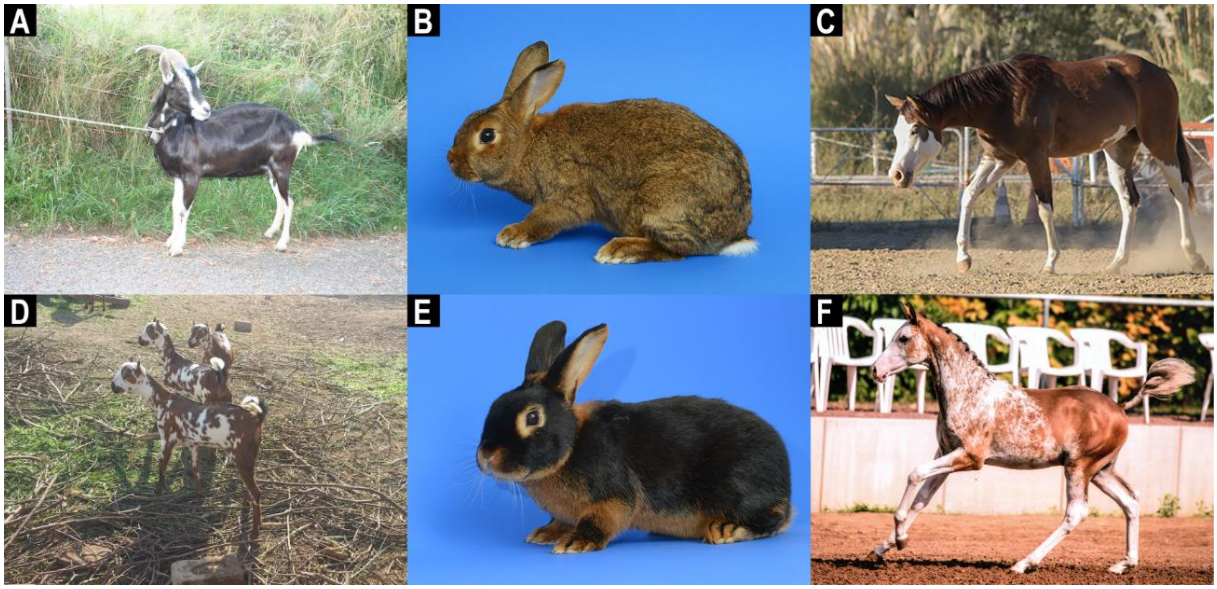

Figure 1 .Various coat color patterns of domestic animals. A Grison striped goat displaying the breed specific pattern. **B** The agouti wild type pattern of rabbits. **C** An American Point Horse with a white spotting coat color phenotype involving depigmentation of the head and the distal extremities. **D** Pakistani Barbari goats showing the breed specific white spotting phenotype. **E** The black and tan coat color of rabbits. **F** A German Riding Pony with a white spotting coat color phenotype involving partial depigmentation on the body and the extremities. Picture B and E are modified from Letko *et al.* 2020 [5], C from Henkel *et al*. 2019 [6] and F from Hug *et al*. 2019 [7]. All images were used with permission.

2.1 Melanocytes – the pigment producing cells

Melanocytes are a heterogeneous group of cells in vertebrates. They are mostly concerned with the production of pigment in skin and hair follicles [8]. Some have other physiological functions than the production of pigment and those are present in the inner ear and heart. Melanocytes produce different forms of melanin in specialized organelles, called melanosomes. Synthesized melanin is deposited in melanin granules and transferred into

keratinocytes and other parts of the skin. These pigment producing melanocytes reside in the basal layer of the epidermis and the bulb of growing hair (Figure 2).

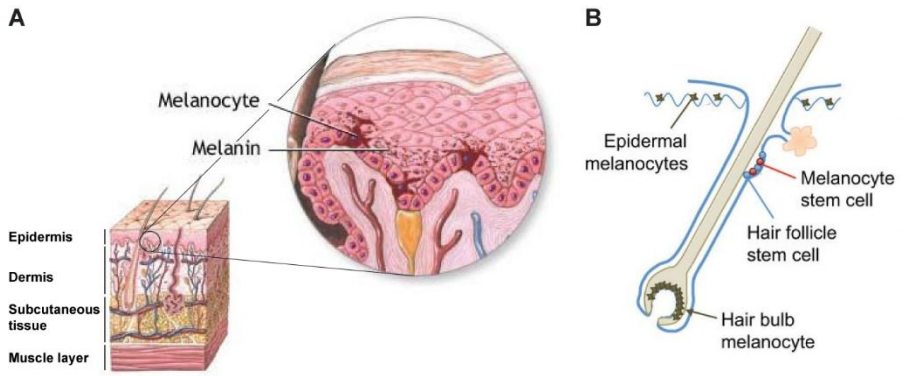

Figure 2. Schematic representation of mammalian skin and hair follicle. A Representation of mammalian skin, separated in four layers (epidermis, dermis, subcutaneous tissue and muscle layer). The insert shows an interfollicular epidermal melanocyte providing pigment to keratinocytes by depositing melanosomes filled with melanin. **B** Melanocyte stem cells (red) reside in the bulge region of the hair follicle and are supported by hair follicle stem cells (blue). Differentiated intrafollicular melanocytes (black) reside at the bulb to provide pigment to the growing hair. Epidermal interfollicular melanocytes are also indicated. Figure A modified from [9] and Figure B from Mort *et al.* 2015 [10], with permission.

2.1.1 Where do they originate?

Melanoblasts are the progenitor cells of melanocytes. Melanocytes are clonally derived from a fixed, small number of primordial melanoblasts in the neural crest region [11,12]. They descend from neural crest cells, a distinct multipotent, migratory, transient stem cell population which gives rise to various cell types early during development, and for that reason are called the "fourth germ layer" [13]. In fact, all pigments cells, except for the cells in the retinal pigment epithelium, originate from neural crest cells [12]. Neural crest cells arise alongside the edges of the neural plate as the neural plate folds, closes to the neural tube and separates from the ectoderm. Thereby all neural crest cells undergo a process called epithelial mesenchymal transition (EMT) followed by delamination to become migratory mesenchymal stem cells. Neural crest cells give rise to a variety of different cell types (e.g. glia cells, adipocytes or melanocytes) depending on their migratory routes and external cues [13] (Figure 3 A). Based on the axial levels of the neural tube, neural crest cells can be grouped into cranial, vagal, trunk and sacral neural crest cells [14], which have their own specific cell fate. While trunk neural crest cells give exclusively rise to melanogenic and neurogenic cells, both cell types are produced along the entire axis [15].

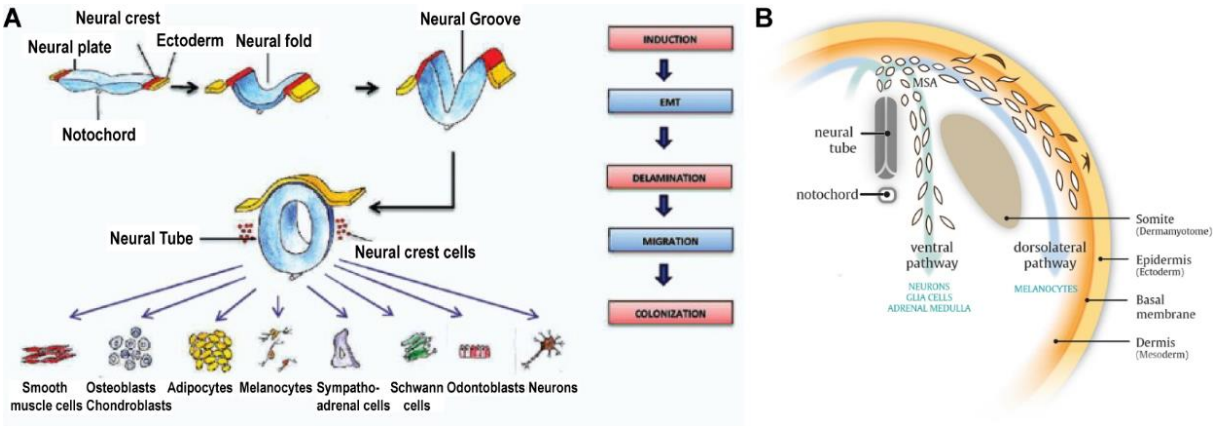

Figure 3. Origin of neural crest cells and model for melanoblast migration. A A model describing different stages in neural crest induction starting with the folding of the neural plate bypassing intermediate stages to the closed neural tube. As the neural tube closes, the neural crest cells undergo epithelial mesenchymal transition (EMT) followed by delamination resulting in multipotent mesenchymal stem cells. Said cells migrate on specific migratory routes to their final destination were they colonize the tissue. During this process, the neural crest cells, proliferate and differentiate into various cell types. **B** Starting from the migratory staging area (MSA) melanocytes differentiated from melanoblasts exclusively via the dorsolateral migration route between the ectoderm and dermomyotome (somite). Neurogenic populations differentiating from ventrally migrating trunk neural crest cells. Ventrally migrating neural crest cells give rise to numerous other lineages, whereas the dorsolateral pathway is restricted to melanoblasts only. Panel A is modified from Shyamala *et al.* 2015 [13], with permission and B adapted from Vandamme and Berx 2019 [15], Springer Nature with permission.

2.1.2 How do they migrate?

In the neural crest region, a middorsal separation exists, so that migration and clonal proliferation of neural crest cells occur autonomously on each side. The migration follows species-specific migration routes which differ significantly in timing and spatial pathways of morphogenetic events [16] as well on the axial level of the neural crest cells. In mammals, after delamination, the neural crest cells accumulate in the migration staging area from which they start migrating into two distinct migration routes, the dorsolateral route and the ventral route (Figure 3 B). The dorsolateral pathway is restricted to melanoblasts [13]. Since neural crest cells are the source of many important cell lines (e.g. melanocytes and smooth muscle cells), the entire process of their delamination, migration and differentiation has to be tightly regulated. Especially as neural crest cells are initially multipotent but lose some of their potential after differentiation [17]. Many studies were conducted on this topic using different vertebrate embryos, such as chicken, mice, zebrafish and xenopus [15]. Said studies helped to shine light on the complex regulatory network of neural crest cell formation were the Wnt signaling molecule (WNTs), bone morphogenetic protein (BMPs), fibroblast growth factors (FGFs) and notch signaling are the main drivers govern the initial steps.

Melanoblasts, as one specific subset of migratory neural crest cells follow the dorsolateral path, towards their target areas in skin and hair. During migration the melanoblasts proliferate and differentiate into melanocytes leading to the distribution of full complements of melanocytes. This process requires another complex regulatory network of signaling

molecules, to govern the developmental program of melanocytes [18]. Figure 4 gives a comprehensive overview of the complex interactions of regulatory genes leading to migratory neural crest cells and the regulation of melanocytes.

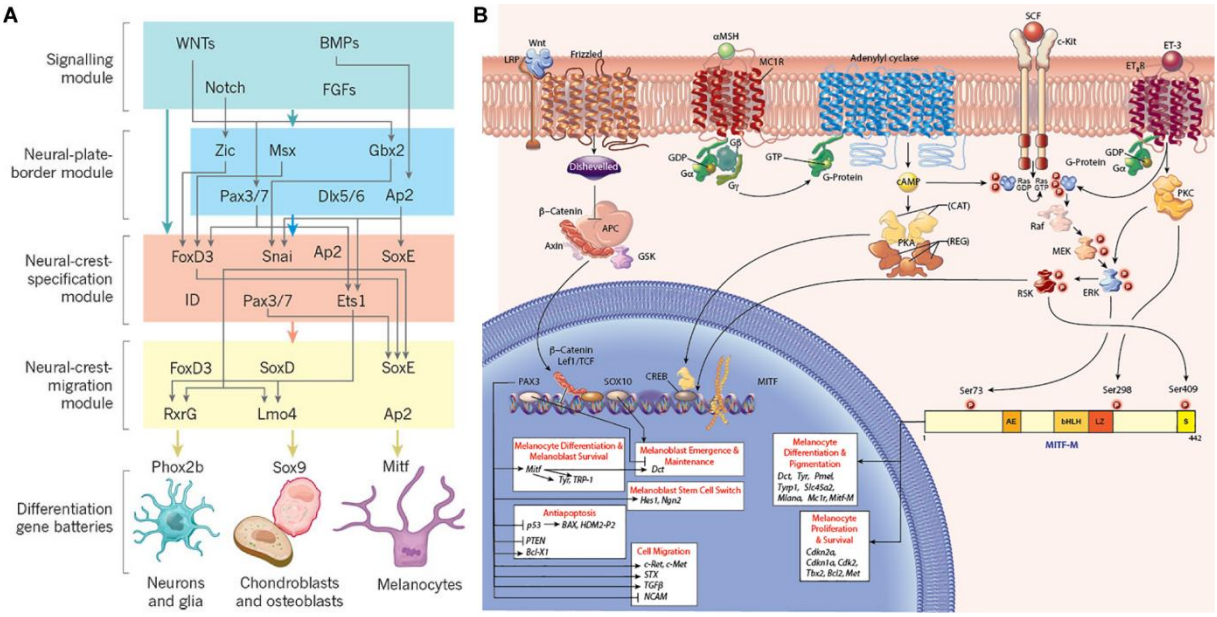

Figure 4. Complex regulatory networks of neural crest formation and melanocyte regulation. A Gene regulation network of neural crest formation in vertebrates. WNTs and BMPs are the main driver and modulates the interaction of different transcription factors controlling neural plate formation, migration and differentiation. The microphthalmia transcription factor (*MITF*) is the main regulator involved in melanocyte development. **B** The gene regulation network of *MITF*, which is central in the formation of melanocytes. The promoter region of the *MITF-M* transcript is shown together with the transcription factors as well as their targets and regulatory effects. Panel A modified from Green *et al.* 2015 [19], Springer Nature with permission. Panel B from [20] with permission.

2.1.3 Key genes of melanocyte development

Genes controlling melanocyte migration, survival and differentiation encode a number of extrinsic signaling factors and their corresponding receptors, as well as downstream signaling molecules such as transcription factors. The key genes of melanocyte development are *KIT, KITL, EDNRB, EDN3, MITF, SOX10, PAX3* and *SNAI2* (Figure 4). It is imperative to discuss the importance of these genes, as variants affecting them give rise to a number of depigmentation phenotypes (Figure 5 A-H), which can be summarized under the term leucism. Leucism describes a phenotype characterized by a reduced number or the complete lack of melanocytes in skin and hair follicles. This has often pleiotropic effects for the individual [21]. The loss of melanocytes results in different forms of white spotting, ranging from few white spots to completely depigmented individuals (Figure 1 C, D, F and Figure 5 A-H), sometimes with additional side effects. Those can be hearing loss, aganglionosis, deficiencies in hematopoiesis and deficiencies in gametogenesis [22–25]. In domestic animals these deleterious side effects have sometimes not been recognized immediately and lead to considerable animal welfare concerns. In humans, the most important depigmentation disorder

is the Waardenburg syndrome (WS), a group of rare genetic disorders. It was first described and named after Petrus Johannes Waardenburg, a Dutch ophthalmologist in 1951 [26]. WS is described as a congenital disorder marked by facial abnormalities like dystopia canthorum, pigmentation deficiencies of hair, skin and iris, and some degree of hearing loss. WS is divided into four general types, which are further subdivided based on the differences in phenotype and severity of symptoms [27]. A common cause of all ten known WS subtypes is, that genes involved in melanocyte development are affected (Table 1). The key genes involved in melanocyte development and important in the formation of WS are discussed in the following.

Table 1. Human pigmentation disorders associated with key genes of melanocyte development. Genes, phenotypes and OMIM entries are given (Online Mendelian Inheritance in Man, https://www.omim.org/).

Gene	Phenotype	OMIM number
KIT	Piebaldism	172800
KITL	Pigmentation abnormalities or deafness	145250, 611664, 616697
EDNRB	WS4A	277580
EDN3	WS4B	613265
MITF	WS2A or WS2C	193510, 606662
SOX10	WS2B, WS2E or WS4C	600193, 611584, 613266
PAX3	WS1 or WS3	193500, 148820
SNAI2	Piebaldism or WS2D	172800, 608890

2.1.3.1 KIT

The KIT receptor tyrosine kinase is conserved across all vertebrates and activated upon binding of its ligand (KITL). KIT is one of the key molecules involved in melanocyte differentiation, migration and survival. With dimerization followed by activation of KIT signaling due to ligand binding, MITF will be regulated by phosphorylation via the MAPK pathway (Figure 4 B, top right, c-KIT) [28]. Besides its important function for melanocytes, KIT is involved in the regulation of hematopoiesis, stem cell maintenance, gametogenesis and in mast cell development [29,30]. Interestingly genetic variants in *KIT* can vary in size and location and still result in a variable depigmentation phenotype (e.g. Figure 5 A). Those can range from subtle white spotting phenotypes towards complete depigmentation in various livestock animals [31–35]. In humans, variants of *KIT* lead to piebaldism, a rare depigmentation phenotype (Table 1) [36]. The complete lack of KIT during early embryonic development is often lethal, further indicating the importance of the gene [37].

2.1.3.2 KITL

KIT ligand, also called stem cell factor (SCF) or steel, is the ligand of KIT leading to its activation to fulfill its important role in melanocyte development (Figure 4 B, top right, SCF) [30]. *KITL* was formerly also termed steel locus (*Sl* in mice) and genetic variants of it give rise to many different phenotypes, including depigmentation phenotypes (Figure 5 B). In humans, mutant KITL can lead to pigmentation abnormalities or deafness (Table 1). Interestingly, KITL is expressed as membrane bound protein with two isoforms. The longer isoform contains a protease cleavage site and will be rapidly processed into soluble dimeric KITL, while the shorter isoform stays as dimer in the plasma membrane [30].

2.1.3.3 EDNRB

Endothelin receptor type B is a seven transmembrane G protein-coupled receptor binding the three vasoactive endothelins (EDN1-3) (Figure 4 B, top right, ET_B-R) [38]. Genetic variation of the gene in humans result in the Waardenburg syndrome (WS4A, Table 1), manifesting in deficient pigmentation and hearing loss due to absent melanocytes and/or Hirschsprung disease characterized by impaired motility of the intestine due to a lack of enteric ganglia [39]. Studies on *ENDRB* in human malignant melanoma suggested its role in the development of melanocytes [40,41] as well as in the enteric nervous system. Mouse experiments indicated that the gene is expressed during early embryonic development in the neural tube and is required for normal melanocyte development [42]. Knockout experiments of *EDNRB* result in animals with extreme depigmentation phenotypes [43] (Figure 5 C), while the paralog of ENDRB, endothelin receptor type A (ENDRA) is not as closely associated with melanocyte development [40,44].

2.1.3.4 EDN3

Endothelin 3 is a vasoactive peptide consisting of 21 amino acid residues. It is expressed in various tissue types but most importantly in neural crest cells and surrounding mesenchymal cells [45]. EDN3 binds specifically to EDNRB and plays, therefore an important role in melanocyte development (Figure 4 B, top right, ET3). Accordingly, genetic variants of *EDN3* in humans result in the Waardenburg syndrome and/or Hirschsprung disease (WS4B, Table1), similar to mutant *EDNRB* [39]. *Edn3* knockout mice have a reduced number of melanoblasts and a white spotting phenotype (Figure 5 D). Remarkably, overexpression of *Edn3* in mice result in a hyperpigmentation phenotype, due to more melanocytes in the dermis [45].

2.1.3.5 MITF

MITF is known as the master regulator of melanocyte development. MITF also regulates the differentiation and development of retinal pigment epithelium and is responsible for melanocyte-specific transcription of the genes encoding enzymes required for melanogenesis (Figure 4 bottom right, M-MITF) [46–49]. MITF belongs to the family of basic helix loop helix domain transcription factors and till date at least nine distinct *MITF* mRNA isoforms are known [50]. All isoforms differ mainly in the first exon while the following eight (exon 2- 9) are usually found in all transcripts. Different isoforms are found to affect various linage specific pathways of different cell types [50]. For melanocyte development the isoform *M-MITF* is essential and exclusively expressed in melanoblasts [47]. MITF is known to directly affect a number of target genes (Figure 4 B) important for melanocyte development and melanogenesis. Variants of the gene lead to various depigmentation phenotypes (Figure 5 E), often associated with a white head, blue eyes and higher risk for deafness [51–53]. Complete loss of function of *MITF* holds the risk of being neonatal or embryonic lethal, indicating the importance of this gene during early embryonic development. Due to its importance, *MITF* is tightly regulated by complex regulatory networks and feed-back loops [47,49,54,55]. In addition to its importance in pigment cell development, *MITF*´s function in eye imaginal disc formation should not be overlooked as it might be conserved across vertebrates and invertebrates [56]. In humans, such variants of *MITF* are associated with two different types of the Waardenburg syndrome (WS2A and WS2C, Table 1), sharing the absence of dystopia canthorum.

2.1.3.6 SOX10

SOX10 belongs to the SRY-related high mobility group (HMG) family of transcription factors known to consist of 20 different proteins in mammals. Their function is well conserved throughout the animal kingdom and they are classified based on function and sequence similarities. SOX10 is considered as the master regulator gene for neural crest development [57]. Together with other SOX proteins (SOX8 and SOX9) it is grouped in the SOXE subgroup important in neural crest and melanocyte development (Figure 4 A) [58,59]. *SOX10* in particular is important in the maintenance of melanoblasts (Figure 4 B) as heterozygous murine mutants display a reduction of melanocytes and consequently depigmentation (Figure 5 F), while homozygous mutants are embryonic lethal [60,61]. In humans, variants of *SOX10* are associated with different forms of Waardenburg syndrome and/or Hirschsprung disease (WS2B, WS2E and WS4C, Table 1) and potentially neurologic abnormalities, including mental impairment, myelination defects, and ataxia [62].

2.1.3.7 PAX3

Paired Box 3 is a transcription factor of the PAX family characterized by a highly conserved DNA binding domain (paired box). It was first described in Drosophila segmentation genes. In mammals, there are nine such genes participating in the regulation of cell migration, proliferation and differentiation [63]. *PAX3* is active in neural crest cells embedded in the complex regulatory network of neural crest formation and melanocyte regulation (Figure 4). It is in particular important in the differentiation of melanoblasts [64]. The 'splotch' mouse mutant is due to a variant in *Pax3* [65]. PAX3 binds to the promotor of *MITF* and activates the transcription of the *MITF-M* mRNA [66]. Genetic variation of the *PAX3* itself as well as variation of the PAX3 binding site in *MITF* result in light white spotting (Figure 5 G) [51,64]. Furthermore, variants of the human *PAX3* gene are associated with different types of Waardenburg syndrome (WS1 and WS3, Table 1), characterized by pigmentation abnormalities and upper limb abnormalities (Klein Waardenburg syndrome, WS3).

2.1.3.8 SNAI2 / SLUG

The Snail-related zing-finger 2 is a transcription factor also known as SLUG or SLUGH, belonging to the Snail superfamily of transcription factors which have an evolutionary conserved role across vertebrates and invertebrates [67]. Members of the Snail superfamily act as repressors [68]. *SNAI2* was first identified in the neural crest of chick embryos and is expressed by cells undergoing EMT (epithelial mesenchymal transition) [69]. Studies in mice show, that *Snai2* is important for melanoblast migration and survival as it is expressed in migratory neural crest cells [70]. Mutants display white spotting of the head and various other body parts (Figure 5 H) indicating the importance of *Snai2* for the normal development of melanocytes [71]. Mutants are viable but display additional severe phenotypes [72]. In humans, genetic variants of *SNAI2* are associated with a type of Waardenburg syndrome (WS2D) similar to that of *MITF* variants and piebaldism similar to *KIT* variants (Table 1).

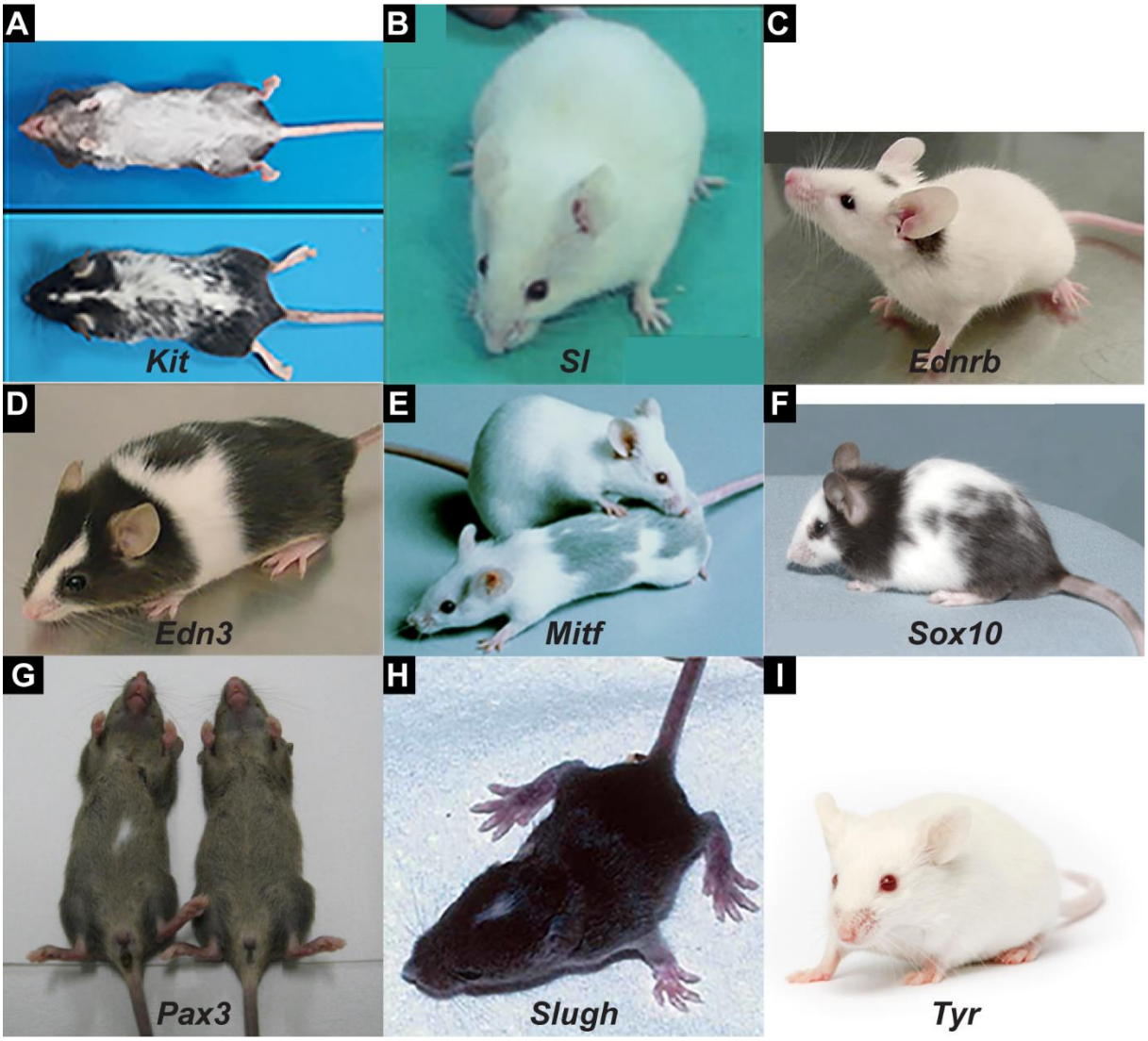

Figure 5. Depigmentation phenotypes in mice. A Example of a depigmentation phenotype caused by a *Kit* variant (*Kit*$^{D17/+}$). Image modified from Sun *et al.* 2020 [73]. **B** Lack of pigmentation of skin in *Sl* (*Kitl*) homozygous mutant mice due to *Sl/Sl*^{d} [25]. Figure modified from Pérez-Losada *et al.* 2002 [72]. **C** White spotting phenotype with only some pigmented spots left caused by a genetic variant of *Ednrb*, known as piebald lethal (*Ednrb*$^{sl/sl}$). **D** A genetic variant of *Edn3* known as lethal spotting (*Edn3*$^{ls/ls}$) causes a white spotting phenotype. Image C and D modified from Saldana-Caboverde and Kos 2010 [74]. **E** Depigmentation phenotype caused by genetic variants in *Mitf*. Homozygous *Mitf* $^{mi\text{-}x39}$ (foreground) and *Mitf* $^{mi\text{-}enu198}$ (background) mice. Image adapted from Hallson *et al.* 2000 [75]. **F** White spotting phenotype of caused by a genetic variant in *Sox10* (*Dalmatian*). Figure adapted from [76]. **G** ventral view of white spotting due to mutant *Pax3* (left) and wild type (right) mice. Image modified from Ohno *et al.* 2017 [77]. **H** White spotting phenotype of a *Slugh* (*Snai2*) -deficient mouse. Image adapted from Sánchez-Martín *et al.* 2003 [78]. **I** Mouse displaying an albinism phenotype caused by a genetic variant in *TYR* (*Tyr*$^{C/C}$). In contrast to the leucistic mouse mutants shown in panels A-H, albinistic mice do not have any pigment in the eyes and consequently have red eyes. Image adapted with permission from © The Jackson Laboratory [79]. All images were used with permission.

2.2 Melanogenesis and pigment-type switching

Melanocytes produce two types of melanin in their melanosomes, the brown to black eumelanin and the yellow to red pheomelanin by a process called melanogenesis [80]. At a given time, a melanocyte will only produce one type of pigment. However, depending on the external stimuli, a melanocyte can chance the synthesized pigment. This so-called pigment-type switching involves binding of one of two signaling molecules α-melanocyte-stimulating hormone (α-MSH) or agouti-signaling protein (ASIP), to the melanocortin 1 receptor (MC1R) (Figure 6 A). Binding influences cellular cAMP levels in the melanocyte leading to the synthesis of eumelanin or pheomelanin. Key genes involved in this synthesis are the enzymes tyrosinase (TYR), dopachroma tauromerase (DCT) and tyrosinase like protein 1 (TYRP1) (Figure 4 B) [81].The regulation of temporal and spatial expression of both melanin types is called pigment-type switching, when binding of either one signaling molecule leads to the expression of exclusively one specific melanin [82–86]. This process is involved in many different coat color phenotypes [83].

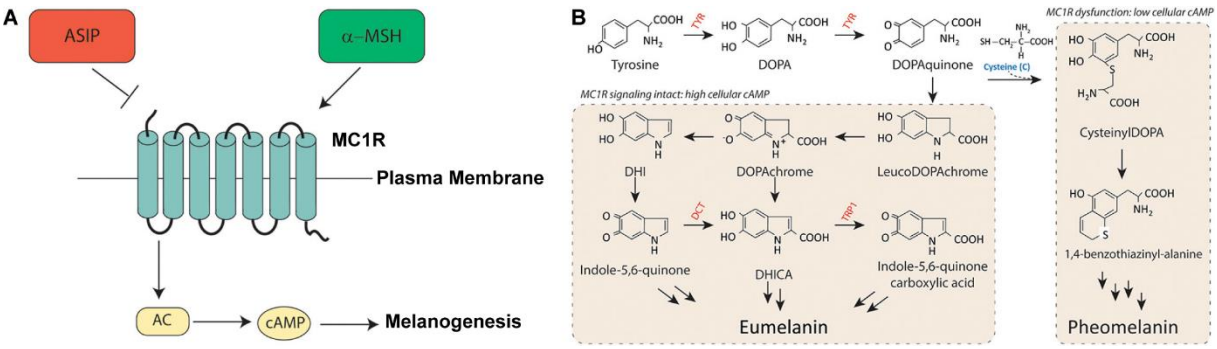

Figure 6. Melanin regulation and production. A Schematic representation of MC1R in the plasma membrane of melanocytes. Binding of α-MSH as the agonist or ASIP as the competitive antagonist, leads to activation or inhibition of MC1R. Activation of MC1R leads to an increase in cellular cAMP levels and the preferred production of eumelanin. Adenylyl cyclase (AC) catalyzes the conversion of ATP to cAMP. **B** Schematic representation of melanin production pathway in melanosomes. Key enzymes are highlighted in red. Panels A and B are modified and adapted from Wolf Horrell *et al.* 2016 [87], with permission.

2.2.1 Key genes of melanogenesis and pigment-type switching

Genes important for melanogenesis and pigment-type switching are mainly two of extrinsic signaling factors and their corresponding receptor, as well as enzymes catalyzing the pigment synthesis. The key genes of pigment-type switching are *MC1R, α-MSH* and *ASIP* [82–84], and for melanogenesis *TYR*, *DCT* and *TYRP1* [88] (Figure 6). Alterations of genes involved in pigment-type switching result in interesting coat color patterns (Figure 1 A), while variants altering melanogenesis, often lead to albinistic individuals (Figure 5 I). Interestingly, a strong positive correlation between MC1R and extensive white markings was observed suggesting a modifier effect of genetic variants of *MC1R* [51,89–91]. This indicates an effect on the number

of melanocyte in skin and hair follicle, despite of genes involved in pigment-type switching and melanogenesis do not directly alter them.

Albinism is an inherited hypopigmentation disorder caused by a partial or complete reduction in melanin biosynthesis in the melanocytes leading to a lack of mature melanin. Most humans with albinism have either oculocutaneous albinism (OCA) or ocular albinism (OA) depending on the affected genes (Table 2) [92]. OCA is inherited autosomal recessive, while OA is X-linked and thereby affecting mostly males. Carriers often have pigmentary mosaicism in the fundi without any side effects. Patients suffering from the most frequent OCA display hypopigmentation of skin, hair and eyes making them susceptible to UV radiation and leading to serious eye dysfunction [93]. OA on the other side is characterized by normal pigmentation of skin and hair follicles, while the eye is depigmented leading to ocular dysfunction. Variants in *GPR143* encoding a G protein-coupled receptor involved in melanogenesis lead to this specific form of albinism [94,95]. Key genes involved in pigment-type-switching and melanogenesis associated with albinism are discussed in the following.

Table 2. Types of oculocutaneous albinism or ocular albinism. Genes associated with the different types and OMIM entries are given (Online Mendelian Inheritance in Man, https://www.omim.org/).

Type	Gene	OMIM
OCA1A	*TYR*	203100
OCA1B	*TYR*	606952
OCA2	*OCA2*	203200
OCA3	*TYRP1*	203290
OCA4	*SLC45A2*	606574
OCA5	unknown	615312
OCA6	*SLC24A5*	113750
OCA7	*LRMDA*	615179
OCA8	*DCT*	619165
OA1	*GPR143*	300500

2.2.1.1 MC1R and α-MSH

Melanocytes produce the yellow to red pheomelanin by default. They switch to the production of eumelanin upon activation of MC1R. MC1R is a G protein-coupled receptor, localized in the plasma membrane of melanocytes with a high affinity for α-MSH and also its precursor, adrenocorticotropin [96,97]. Binding of the agonist α-MSH induces a conformation change of MC1R that activates the enzyme adenylyl cyclase (AC) by signal transmission via the G protein. Then AC catalyzes the conversion of ATP to cAMP, resulting in elevated cellular cAMP levels triggering the switch of pheomelanin to eumelanin production in the melanosomes (Figure 6).

2.2.1.2 ASIP

ASIP is the competitive antagonist of α-MSH and can also bind to MC1R [98–100]. Binding of ASIP blocks binding of α-MSH, inhibiting the activation of MC1R, repressing AC activation and thereby reducing cellular cAMP levels. When cAMP levels are low, pheomelanin is produced in the melanosomes (Figure 6).

2.2.1.3 TYR

Tyrosinase (TYR) is an enzyme catalyzing the first two initial and rate limiting steps of melanin synthesis. It first mediates the hydroxylation of tyrosin to dihydroxyphenylalanine (DOPA) and then immediately catalyzes the oxidation to DOPAquinone (Figure 6 B). Genetic variants of *TYR* lead to albinistic individuals (Figure 5 I) and in humans causes two types of OCA (OCA1A and OCA1B, Table 2) [101].

2.2.1.4 DCT

L-dopachrome tautomerase (DCT), also known as tyrosinase-related protein 2 (TYRP2) is another key enzyme in melanin synthesis, especially for eumelanin synthesis. It catalyzes the conversion of L-DOPAchrom to DHI-2-Carboxylic acid (DHICA) (Firgure 6 B) [102]. In humans, variants of *DCT* cause one form of OCA (OCA8, Table 2) [103]. Historically *DCT* was established as an early melanoblast marker [104] used in various studies dissecting melanoblast development [55].

2.2.1.5 TYRP1

Tyrosinase-related protein 1 (TYRP1) plays an important role in eumelanin synthesis. It catalyzes the transformation of DHICA into indole-5,6-quinone-2-carboxylic acid (Figure 6 B) [105,106]. Genetic variants of *TYRP1* causing one form of OCA in humans (OCA3, Table 2) [107,108]. In domestic animals and mice variants lead to alterations in coat color [109–112].

2.3 Domestic animals as model organisms for coat color phenotypes

The process of domestication involves adaptation of wild species to the environment man provides. Adaptation is achieved through genetic change over generations to environmental influence [113]. Domestication involves a process called artificial selection guided by their owners [114]. In contrast to artificial selection, natural selection is a non-intentional process by adapting to changing environments which introduce selective changes to the individual [115–117]. During early domestication, humans almost exclusively selected animals for breeding based on behavioral characteristics [118]. Physiological or morphological alterations were often secondary byproducts of this selective pressure [119]. Since the initial process of domestication, humans then shaped domestic animals based on favorable attributes such as fur, food, or transportation [120]. This strong artificial selection of domestic animals is an evolutionary process linked with changes in the gene pool resulting in an enrichment of gene variants associated with these favorable phenotypes [121]. As a consequence of the long period of strong phenotypic selection, distinctive breeds were formed which are often kept as closed populations. These breeds are characterized by a specific set of phenotypes and typically the most distinctive ones for differentiation are related to morphology or coat color as they are easy to observe [121]. Due to the abundance and relatively short reproduction cycle (compared to humans), domestic animals are well suited to observe spontaneous mutations resulting in interesting phenotypes [122]. Such changes can be either harmless or harmful to the individual. As for coat color, an interesting variation of the breeds standard coat color might evolve, which is of particular interest to owners and breeders and therefore, will be kept in the population, rising in frequency [3]. The investigation of these specific coat color phenotypes gives us the opportunity to understand not only the underlying molecular mechanisms of coat color itself, but as well giving us insight into the breed's history [122–124]. This makes domestic animals a well suited model for the dissection of various pathways of interest as studies conducted on domestic animals give a comprehensive view on genotype-phenotype relationships [121,122].

2.4 Selected methods and technologies in genetic analyses

In the beginning, the investigation of the genetic causes for heritable traits was a challenging task relying on candidate gene approaches. This not only required biological understanding of the phenotype and therewith a set of candidate genes but also capable technical methods. The genetic analyses as we know it today, began with DNA sequencing in the 1970s [125] and advances with Sanger sequencing [126]. It made sequencing of candidate genes for a particular phenotype possible, potentially revealing the underlying genetic cause. Another important breakthrough toke place in 1983 with the invention of the polymerase chain reaction

(PCR) by Mullis [127], further facilitating targeted sequencing using PCR and Sanger sequencing. In the 1990s, positional cloning became popular to locate the chromosomal position of disease associated genes. It even works without knowledge or hypotheses about the biological basis of the disorder, which made it a dominant method for disease gene discovery [128]. Two important widely applied methods comprised in positional cloning are linkage analysis [129] and homozygosity mapping [130].

All of the used methods require some form of genotype data with corresponding chromosomal positions. Today, genotype datasets are often made up of thousands of single nucleotide variants (SNVs) produced by either species specific genotyping arrays [131] varying in marker density or next generation sequencing.

2.4.1 Next generation sequencing (NGS)

The first generation of sequencing technique started with the initial invention of DNA sequencing [125] and advances with Sanger sequencing [126]. This first generation sequencing technique is still applied today for targeted sequencing of candidate genes and validation of next generation sequencing results. The reason is, that it produces high accurate results of around one kilobase in length for a relatively low experimental price [132].

With the development of next generation sequencing techniques (second and third generation) high-throughput sequencing became possible. This came with a reduction in sequencing cost, time and effort [133] but often required advanced computational power to handle the produced big datasets [134]. These sequencing methods revolutionized genomics, as DNA and RNA sequencing became feasible to everyone, enabling its use for research and diagnostic purposes [132,135,136]. The different developed NGS platforms differ not only in the principles of their methodology but also in other factors such as speed, read length and throughput [137]. Nowadays, the by far most commonly used NGS technology of the second generation, is Illumina followed by Ion Torrent [138].

Illumina machines are short read sequencing devices and they most commonly produce single or paired end reads of 150 bp in length. The use of bioinformatics tools to analyze these short read sequencing data reveal a large portion of the genetic variation, namely the SNVs and short indels [139]. The number of identified SNVs typically exceed those of genotyping arrays by a large margin (Affymetrix Genome-Wide Human SNP Array 6.0 = 1.8 million genotypes vs ~20 million SNVs per human genome [140]). This large SNV dataset represents a part of the normal genetic variation of the sequenced individual. The number might vary from species to species due to species-specific parameters and the quality level of the available reference genome. If the sequenced individual had a rare genetic disorder caused by a single genetic variant, the causative variant will be present in this dataset. It is possible to identify causative variants in a relatively small number of cases and controls by comparing their variant datasets

with each other [141]. Particularly for monogenic disorders caused by rare variants, this approach of using NGS technologies has had a huge impact and has led to a fast increase in the number of known causative variants [130,133].

One major drawback of the common second generation sequencing technologies is the short read sequencing with the consequence of not been able to detect the whole genetic variation of an individual, as large structural variants are often not detectable [135]. Third generation sequencing technologies with much longer read lengths facilitate the analysis of structural variants [136].

The two main protagonists of third generation sequencing technologies are the single molecule real time (SMRT) platform of Pacific Biosciences [142] and the nanopore sequencing [143,144] of Oxford Nanopore. Both platforms produce long to ultra-long reads enabling not only structural variant detection but also *de novo* genome assemblies [142,145].

For this thesis I obtained many whole genome sequencing datasets of different species, exclusively produced on Illumina instruments. I used these datasets not only for the calculation of selection signatures but also for the discovery of functional variants causing heritable coat color phenotypes.

2.4.2 Genetic variation

Genetic variation describes the differences in DNA sequences within and between individuals of the same species. This variation is caused by mutations and is heritable, if mutation events affect the germ line. Genetic variants may be classified based on their size and/or their functional effects. Genetic variants can have multiple effects, from altering gene function by changing the coding sequence and therewith the protein sequence, to non-coding variants with no clear impact. The most impactful chances altering gene function include start/stop codon alterations and frameshift mutations followed by non-synonymous amino acid changes. Such alterations are often straight forward to detect and to assess but they do not display the majority of the variant catalogue. The evaluation of the potential impact of noncoding variants is challenging and requires a more advance experimental setup [146,147].

2.4.2.1 SNVs and short indels

The most commonly observed genetic variants are single nucleotide variants (SNVs), in which one nucleotide is substituted by another. Short indels are insertions or deletions of DNA affecting up to 50 nucleotides. Using short read sequencing, those are the two main types of genetic variation detected and reported in the variant call format (VCF) [139]. One state of the art tool to detect SNVs and short indels is the HaplotypeCaller of the Genome Analysis Toolkit (GATK) [148].

2.4.2.2 Structural variants

Besides SNVs and short indels, structural variants (SVs) play a prominent and important role for not only phenotypic evolution but also for disease associated traits [140]. SVs include duplications, deletions, inversions, translocations and copy number variants (CNV), spanning more than 50 bp [121]. Larger SVs are generally more difficult to detect using short read data as they are not reported in the SNV and short indel dataset [149]. Besides available structural variation calling tools (e.g. BreakDancer [150] or CNVfinder [151]), the most reliable approach to detect them is the visual inspection of the short read alignments of limited genomic regions as it was done for projects in this thesis. Another way would be the usage of third generation sequencing technologies paired with novel SV calling tools (e.g. SMRT-SV [152] or NextSV [153]) [154].
A common pattern for structural changes discovered so far, is that they may lead to an altered configuration of regulatory elements by duplication [35] ,deletion [155] or translocating [156–158] regulatory elements in relation to the coding sequence and thereby alter gene expression.

2.4.3 Methods for identifying causal gene variants

In order to identify causative variants, commonly used methods are based on either one of the two principles: hypothesis- free or hypothesis-driven research.
Hypothesis-free research is the principle of using unbiased holistic tools to spot differences between affected and unaffected individuals without mechanistic understanding of the investigated phenotype [159]. Important methods used for hypothesis-free approaches are "Linkage analysis" (2.4.3.1), "Homozygosity mapping" (2.4.3.2), "Genome wide association study" (2.4.3.3) and "Selection signature" analysis (2.4.3.4), which are outlined in this chapter (2.4.3).
Hypothesis-driven research is the principle of developing a (working) hypothesis about the potential cause for a heritable trait. The developed hypothesis can be experimentally tested and ultimately be true or false. Hypothesis-driven approaches in genetics require some degree of mechanistic understanding how heritable phenotypes are controlled. The most important hypothesis-driven method in forward genetics is the candidate gene approach (2.4.3.5) displayed in greater detail in this chapter (2.4.3).

2.4.3.1 Linkage analysis

Linkage analysis is a method to detect chromosomal regions inherited together in a given family, displaying a trait of interest [160]. It can be performed either non-parametric (model-free) or parametric depending on the complexity of the investigate disorder. Important for linkage analysis are the principles of linkage and recombination [161].

2.4.3.2 Homozygosity mapping

Homozygosity mapping, also called autozygiosity mapping, is a powerful method to locate causal variants, under the premise of being identical by descent. The method works for (monogenic) autosomal recessive traits as it is based on the assumption that an affected individual has inherited two identical copies of chromosomal regions carrying the disease allele from an ancestor, to which it is inbred. Extended homozygous regions are known as runs of homozygosity (ROHs), ranging from usually few hundred kilobases to several mega-bases in size. Their length decreases with genetic distance due to recombination events [130].

2.4.3.3 Genome wide association study (GWAS)

In contrast to the family based methods (e.g. linkage analysis), GWASs are a popular alternative. The classical design of a GWAS is a case-control study where the association of genetic markers of unrelated individuals with a trait is calculated [162]. If the frequency of an allele is significantly different between the case and control group, it is associated with the trait of interest. Most performed GWASs use markers of genotyping arrays as input, meaning that an associated marker is most likely not causative but in linkage disequilibrium with the true causative variant [162].

2.4.3.4 Selection signatures

Selection leaves unique genetic patterns or footprints in the genome behind. These patterns are called selection signatures, representing reduced regional variation surrounding the selected and fixed beneficial alleles [116,163]. In modern times, one of the key interests of animal genetics is to investigate regions under selection in order to identify the underlying cause of a given phenotype and to learn something of their evolutionary history [164]. Therefore, methods based on linkage disequilibrium, allele frequency spectrum, reduced local variability, and haplotype characteristics (Figure 7), are applied to SNV datasets of multiple individuals produced by either genotyping arrays or NGS [165]. One way to use those methods is to search for domestication genes in livestock as the group of Leif Andersson did. They calculated pooled heterozygosity scores (H_p) for chicken, pigs, rabbit and herring [35,166–168]. For this thesis, I chose the same approach for goats in order to identify regions under selection associated with various coat color phenotypes. Additionally, I combined in my study H_p scores with the calculation of fixation index values (F_{ST}) [169], another method to detect selection signatures.

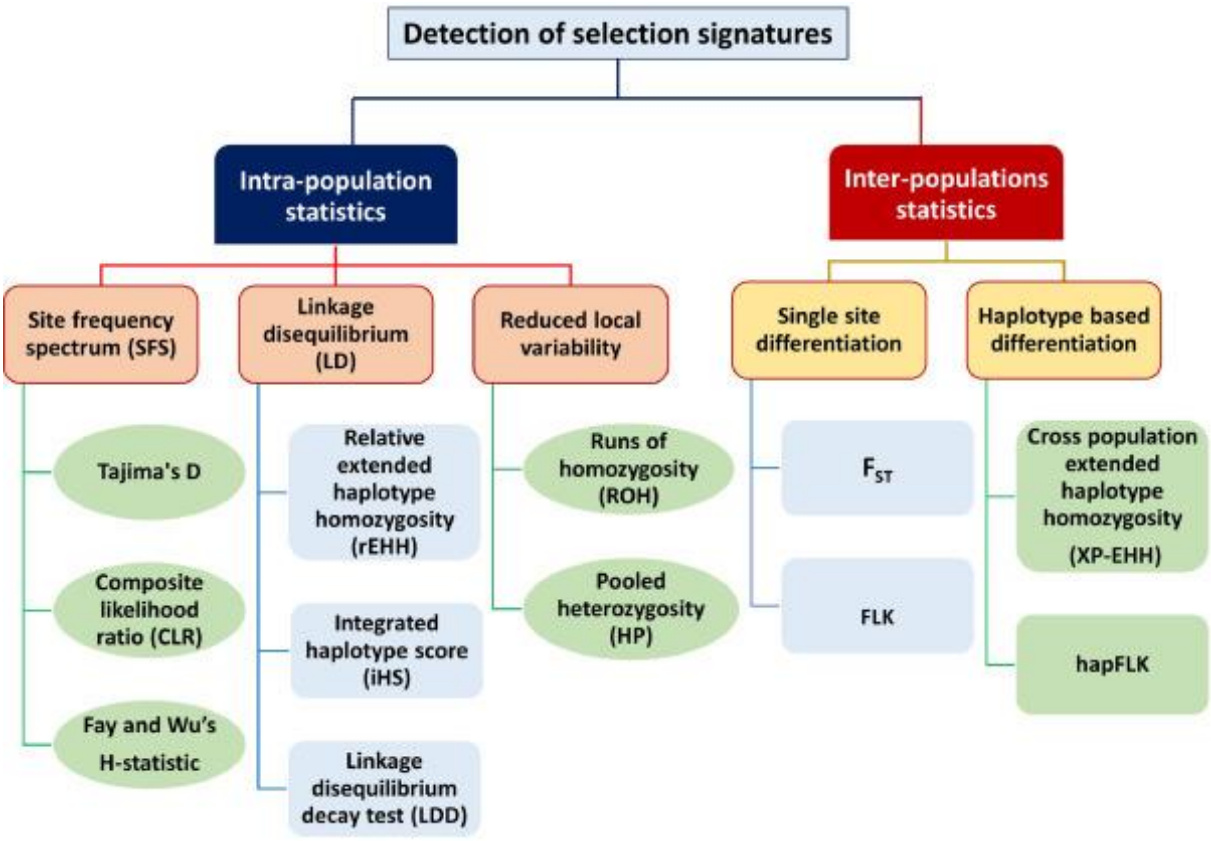

Figure 7. Classification of different approaches commonly used for the detection of selection signatures in livestock populations. Image adapted from Saravanan *et al.* 2020 [165], with permission.

2.4.3.5 Candidate gene approach

The advances of whole genome sequencing (WGS) and the associated reduction of cost, made it feasible to have comprehensive SNV datasets of case and control sequences. In order to identify causal variants for heritable traits in WGS datasets, mapping of the obtained sequenced reads to a reference genome sequence and calling variants are required [170]. Typically several million variants are called compared to the reference genome [139,140]. To identify a causal variant for a given trait, an approach has to be applied to reduce the immense number of variants to a manageable amount. Usually the approach involves hierarchical filtering strategies and variant annotation followed by candidate gene filtration. Modern annotation tools [171] suggest the potential impact of called variants. Filtering for variants with potentially altering impact in a candidate gene set often reveals candidate causative variants. Finally, these variants have to be validated leading hopefully to one true causative variant. This approach works usually fine for rare monogenic diseases and was applied for projects discussed in this thesis.

3. Aim and hypothesis of thesis

Hypothesis

For my thesis I had the following hypotheses: If a breed specific coat color phenotype is investigated, the causative variant has to be under selection and a particular allele is therefore fixed in that particular breed. This means on the other side, it has to be absent (or rare) in other unrelated breeds not showing the same phenotype. Furthermore, if an inheritable, abnormal coat color phenotype (depigmentation phenotype) is investigated, the causative variant has to be present in the affected individual/family. Meanwhile, the variant has to be absent (if inherited as a dominate trait) or in heterozygous state (if inherited as a recessive trait) in unaffected individuals of the same breed or species. The causative variant has to be present in all affected individuals displaying the same abnormal phenotype.

Aim

The aim of this thesis was to identify genetic causative variants in the following domestic mammalian species displaying interesting heritable coat color phenotypes.

1. Pakistani goat breeds with breed-specific white spotting or depigmentation phenotypes
2. Swiss goat breeds displaying interesting coat color patterns
3. Rare coat color phenotypes segregating in Swiss Valais Blackneck goats
4. The black and tan coat color phenotype of rabbits
5. White spotting phenotype of a German Riding Pony
6. American Paint horse family with diverse depigmentation phenotypes and an increased risk of deafness

4. Results

4.1 Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes

Journal:	PLoS Genetics
Manuscript status:	published
Contributions:	genetic analyses, illustrations, original draft, review and editing of manuscript
Version:	Published version
DOI:	10.1371/journal.pgen.1008536

PLOS GENETICS

RESEARCH ARTICLE

Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes

Jan Henkel[1,2], Rashid Saif[1,3], Vidhya Jagannathan[1,2], Corinne Schmocker[1], Flurina Zeindler[4], Erika Bangerter[5], Ursula Herren[5], Dimitris Posantzis[6], Zafer Bulut[7], Philippe Ammann[8], Cord Drögemüller[1,2], Christine Flury[4], Tosso Leeb[1,2]*

1 Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland, **2** DermFocus, University of Bern, Bern, Switzerland, **3** Institute of Biotechnology, Gulab Devi Educational Complex, Lahore, Pakistan, **4** School of Agricultural, Forest and Food Sciences, Bern University of Applied Sciences, Zollikofen, Switzerland, **5** Swiss Goat Breeding Association, Zollikofen, Switzerland, **6** Attica Zoological Park, Spata, Greece, **7** Department of Biochemistry, Faculty of Veterinary Medicine, Selcuk University, Konya, Turkey, **8** ProSpecieRara, Basel, Switzerland

* Tosso.Leeb@vetsuisse.unibe.ch

Citation: Henkel J, Saif R, Jagannathan V, Schmocker C, Zeindler F, Bangerter E, et al. (2019) Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. PLoS Genet 15(12): e1008536. https://doi.org/10.1371/journal.pgen.1008536

Editor: Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, UNITED STATES

Received: August 12, 2019

Accepted: November 23, 2019

Published: December 16, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pgen.1008536

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Abstract

Domestication and human selection have formed diverse goat breeds with characteristic phenotypes. This process correlated with the fixation of causative genetic variants controlling breed-specific traits within regions of reduced genetic diversity, so called selection signatures or selective sweeps. Using whole genome sequencing of DNA pools (pool-seq) from 20 genetically diverse modern goat breeds and bezoars, we identified 2,239 putative selection signatures. In two Pakistani goat breeds, Pak Angora and Barbari, we found selection signatures in a region harboring *KIT*, a gene involved in melanoblast development, migration, and survival. The search for candidate causative variants responsible for these selective sweeps revealed two different copy number variants (CNVs) downstream of *KIT* that were exclusively present in white Pak Angora and white-spotted Barbari goats. Several Swiss goat breeds selected for specific coat colors showed selection signatures at the *ASIP* locus encoding the agouti signaling protein. Analysis of these selective sweeps revealed four different CNVs associated with the white or tan (A^{Wt}), Swiss markings (A^{sm}), badgerface (A^{b}), and the newly proposed peacock (A^{pc}) allele. RNA-seq analyses on skin samples from goats with the different CNV alleles suggest that the identified structural variants lead to an altered expression of *ASIP* between eumelanistic and pheomelanistic body areas. Our study yields novel insights into the genetic control of pigmentation by identifying six functionally relevant CNVs. It illustrates how structural changes of the genome have contributed to phenotypic evolution in domestic goats.

Author summary

Domestic animals have been selected for hundreds or sometimes even thousands of years for traits that were appreciated by their human owners. This process correlated with the

Funding: This study was funded by a grant from the Swiss National Science Foundation (31003A_172964). R.S. was supported by a Swiss Government Excellence Scholarship and a supplementary grant from the Hans Sigrist Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

fixation of causative genetic variants controlling breed-specific traits within regions of reduced genetic diversity, so called selection signatures or selective sweeps. We conducted a comprehensive screen for selection signatures in 20 phenotypically and genetically diverse modern goat breeds and identified a total of 2,239 putative selection signatures in our dataset. Follow-up experiments on selection signatures harboring known candidate genes for coat color revealed six different copy number variants (CNVs). Two of these CNVs were located in the 3'-flanking region of *KIT* and associated with a completely white coat color phenotype in Pak Angora goats and a white-spotted coat color phenotype in Barbari goats, respectively. The other four CNVs were located at the *ASIP* locus. They were associated with four different types of coat color patterning in seven Swiss goat breeds. Their functional effect is mediated by region-specific quantitative changes in *ASIP* mRNA expression. Our study illustrates how structural changes of the genome have contributed to phenotypic evolution in domestic goats.

Introduction

Goat domestication started around 10,000 years ago in the fertile crescent and is believed to be one of the earliest domestication events of livestock animals [1, 2]. Bezoars, the wild ancestors of domestic goats are an extant species with a distribution in Western Asia from Turkey to Pakistan. Since domestication, goats followed the human migration [3] and played an economically important role for their owners by providing various products like milk, meat or fibers. These economical values were further increased by production-orientated breeding, which led to more than 600 diverse goat breeds at present time [4–6].

Artificial selection of domesticated goats not only resulted in specialized elite breeds for milk, meat or fibers, but also in breeds with unique coat color phenotypes [4, 7]. Due to their striking appearance, these goat breeds are of special value to their owners, selected for uniform coat color, and kept in closed populations. Coat color phenotypes are one of the most intensively studied traits in goats [8–12]. They include solid colored animals of different color, animals with symmetrical color patterns, and animals with white markings, white spotting phenotypes or completely white animals.

White markings, white spotting and completely white phenotypes typically result from a lack of melanocytes in the skin and hair follicles. This group of phenotypes is also termed leucism or piebaldism and characterized by defects in melanoblast development or migration [13–17].

Very light coat colors resembling white are also seen in animals that have a normal set of melanocytes synthesizing a very pale pheomelanin [18]. Melanocytes produce two types of pigments, the brown to black eumelanin and the red to yellow pheomelanin. The so-called pigment type switching, an intensively studied signaling process, governs whether a given melanocyte produces eumelanin or pheomelanin [19]. Eumelanin is produced, if MC1R is activated by its ligand α-melanocyte stimulating hormone (α-MSH), while pheomelanin is produced if α-MSH is absent and/or outcompeted by binding of the competitive antagonist ASIP to MC1R [20–25]. Different alternative promoters of the *ASIP* gene enable spatially and temporally regulated *ASIP* expression, which results in characteristic patterns of eumelanin and pheomelanin synthesis [25–28].

Domestication and artificial selection correlated with the fixation of causative genetic variants controlling breed-specific traits within regions of reduced genetic diversity, so called selection signatures or selective sweeps [29–31]. A method detecting regions of low

heterozygosity from sequence data of pooled individuals (pool-seq) was developed and used to identify loci under selection in chicken, pigs, rabbits and the Atlantic herring [32–35]. Recently, pooled heterozygosity scores were also applied to monitor loci under selection in goats [36].

In the present study, we aimed to gain a better understanding of the genetic variants determining breed-specific coat color phenotypes in goats. We therefore performed a comprehensive screen for selection signatures in bezoars and 20 breeds of domesticated goats.

Results

Selection signature analysis

For the present study, we selected 8–12 animals each from 20 phenotypically diverse domesticated goat breeds and their wild ancestor, the bezoar. We isolated genomic DNA from these animals and prepared equimolar DNA pools for sequencing (Table 1).

We obtained 2x150 bp paired-end sequence data corresponding to 30x genome coverage per pool, called high confidence SNVs and calculated pooled heterozygosity scores ($-ZH_p$) in 150 kb sliding windows with 75 kb step size (S1 Fig; S1 and S2 Tables). The significance threshold was conservatively set at $-ZH_p \geq 4$, which identified 5,220 windows with extremely reduced heterozygosity (0.8% of all windows). Overlapping windows were further merged into 2,239 selection signatures (1.1% of total genomic length). This corresponded to 112 selection signatures per breed pool on average (median = 81; S3 Table).

To evaluate the validity of the pool-seq approach, we compared the results from pool-seq data to individual whole genome sequence data of 120 goats from five different Swiss breeds (S1 Table). We called SNVs from the individual sequence data and calculated H_p and $-ZH_p$

Table 1. Breeds collected for pool sequencing (pool-seq).

Breed	Abbreviation	Breed origin	Animals per pool
Pak Angora	ANG	Pakistan	10
Appenzell goat	APZ	Switzerland	12
Barbari	BAR	Pakistan	12
Beetal	BEE	Pakistan	12
Bezoar (*Capra aegagrus*)	BEZ	wild ancestor	8
Grisons Striped goat	BST	Switzerland	12
Boer goat	BUR	Africa	12
Capra Grigia	CAG	Switzerland	12
Dera Din Panah	DDP	Pakistan	12
Chamois Colored goat	GFG	Switzerland	12
Kamori	KAM	Pakistan	12
Nachi	NAC	Pakistan	12
Nera Verzasca	NER	Switzerland	12
Pahari	PAH	Pakistan	12
Peacock goat	PFA	Switzerland	12
Saanen goat	SAN	Switzerland	12
St. Gallen Booted goat	STG	Switzerland	10
Teddy	TED	Pakistan	12
Toggenburg goat	TOG	Switzerland	12
Valais Blackneck goat	VAG	Switzerland	12
African Dwarf goat	ZWZ	Africa	12

https://doi.org/10.1371/journal.pgen.1008536.t001

scores respectively (S2 Table). The pool-seq dataset and the dataset with individual sequences yielded similar results (S2 Fig).

As a validation of the significance threshold, we inspected our data for selection signatures near known causative variants for breed-defining coat color traits. The characteristic brown coat color of the Toggenburg goat is caused by a missense SNV in the *TYRP1* gene, p.Gly496Asp [10]. Toggenburg goats showed the expected selection signature harboring the *TYRP1* gene with a $-ZH_p$ value of 4.88 (Fig 1; S3 Table).

In addition to the search for reduced heterozygosity, we calculated F_{ST} values for each breed pool in a pairwise comparison with bezoars. The F_{ST} analysis identified 847 selection signatures or 0.4% of the total genomic length (S1 and S3 Figs, S4 Table).

CNVs at the *KIT* locus in two Pakistani goat breeds

The completely white Pak Angora and the white spotted Barbari breeds showed strong selection signatures harboring the *KIT* gene on chromosome 6 with $-ZH_p$ values of 7.20 and 4.56 (Fig 1; S3 Table). We searched for candidate causative variants within the signatures, but did not detect any coding variants in the *KIT* gene. However, visual inspection of the short read alignments revealed two different copy number variants (CNVs) downstream of the *KIT* gene in the Pak Angora and Barbari breeds (Fig 2).

Both CNVs started ~63 kb downstream of *KIT* and covered ~100 kb of the genome reference sequence without known coding DNA. The short read-aligments of read-pairs spanning the amplification breakpoints confirmed that the individual copies of the CNV were arranged in tandem in a head to tail orientation (S4 Fig). The Pak Angora allele consisted of a triplication of the 100 kb region. The Barbari allele represented a duplication of the same ~100 kb region with an additional 16,280 bp deletion in its central part. The read-pair information at the breakpoints indicated that the deleted part was replaced by a 22,702 bp genomic fragment from the 5'-flanking sequence of the *RASSF6* gene, which is located 19 Mb further downstream on the same chromosome (S4 Fig; S5 Table). The shared breakpoints of the two different CNV alleles suggested a common origin of these alleles.

CNVs at the *ASIP* locus in Swiss goat breeds

Five Swiss goat breeds with different coat color patterns had a selection signature with $-ZHp \geq 4$ in the region of the *ASIP* locus on chromosome 13 (Fig 1; S3 Table). We did not find any *ASIP* coding variation in the breeds with *ASIP* selection signatures.

Based on the segregation of coat color patterns in a large breeding experiment, the existence of up to 11 different caprine *ASIP* alleles has been postulated [8]. The most dominant of this allelic series, termed "white or tan" (A^{Wt}), is responsible for white coat color in goats [8]. Furthermore, it has been shown that the white coat color in Saanen goats is caused by a triplication of the *ASIP* gene [9].

Inspection of the short-read alignments of our own sequence data from Saanen goats at the *ASIP* locus confirmed the previously reported triplication and revealed the exact boundaries of the triplication. It spans 154,677 bp of the reference genome sequence and comprises the entire coding sequence of the *ASIP*, *AHCY* and *ITCH* genes. The individual copies are arranged in tandem in a head to tail orientation. Appenzell goats, another white goat breed, had the same CNV allele as the Saanen goats (Fig 3, S4 Fig, S5 Table).

We next investigated the coverage plots of Grisons Striped goats and Toggenburg goats. These two breeds show a characteristic color pattern, which has been postulated to be caused by an *ASIP* allele termed "Swiss markings" (A^{sm}) [8]. They are fixed for an *ASIP* allele with 8

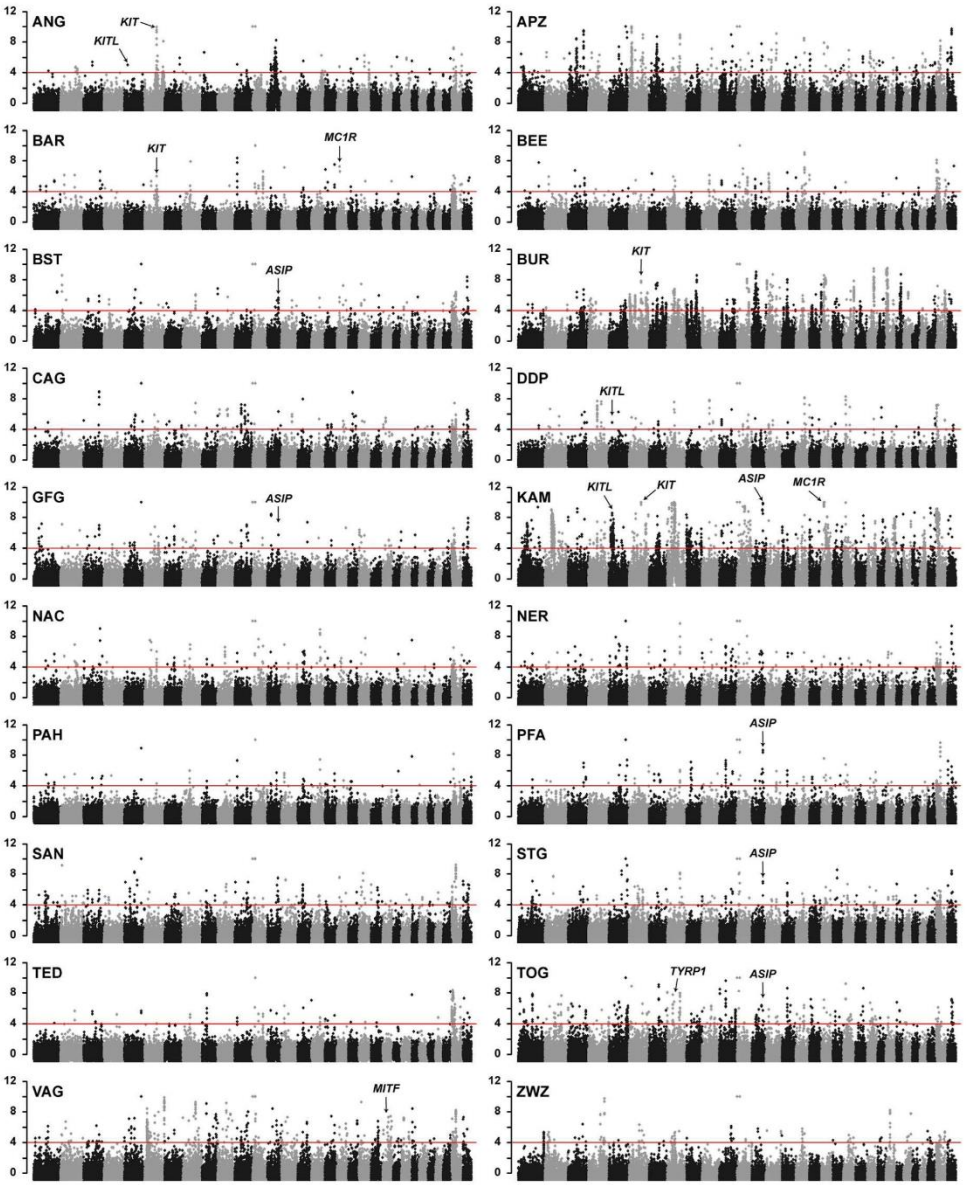

Fig 1. Manhattan plots showing –ZHp values from 20 diverse goat breeds. The red horizontal line indicates the chosen significance threshold of –ZHp = 4. Each dot represents a 150 kb window. Each plot contains 29 autosomes and two unplaced scaffolds representing the X chromosome. Selection signatures co-localizing with known coat color genes are marked with arrows.

https://doi.org/10.1371/journal.pgen.1008536.g001

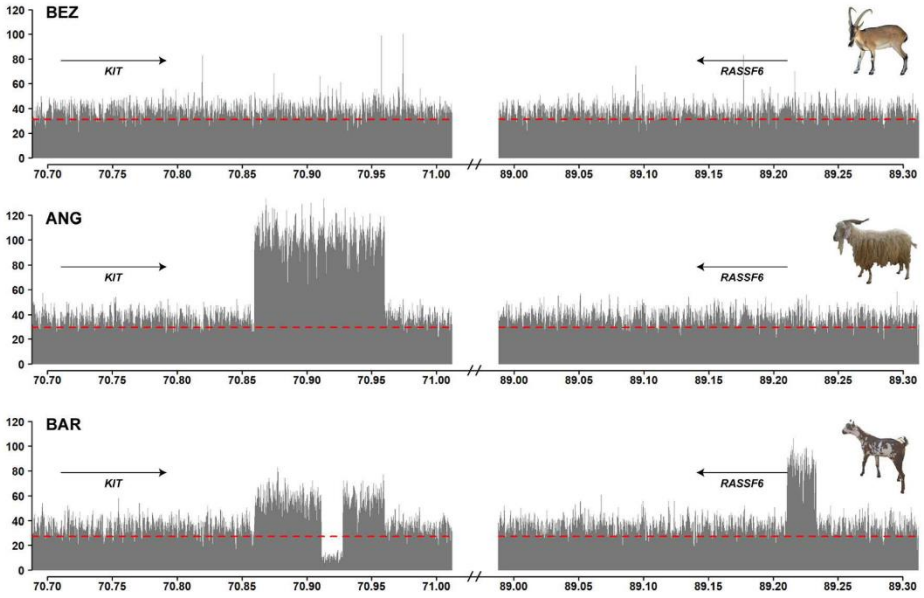

Fig 2. CNVs at the *KIT* locus. The coverage plot of bezoars (BEZ) does not show any copy number variation and represents the wildtype allele. In the Pak Angora breed (ANG), the coverage plot shows a triplication of ~100 kb downstream of the *KIT* gene. In the Barbari breed (BAR), the same region is duplicated. The Barbari allele shows a complex rearrangement involving the insertion of a ~23 kb genome segment originating at 89.2 Mb into the duplicated sequence at ~70.9 Mb with the simultaneous deletion of ~16 kb of *KIT* sequence. Please note that the coverage at ~89.2 Mb corresponds to three times the average. One genome equivalent corresponds to the wildtype sequence at ~89.2 Mb. Read-pair information indicated that the other two genome equivalents are inserted into the duplicated sequence at ~70.9 Mb (S4 Fig). The dashed red line indicates the average coverage across the whole genome of each pool-seq dataset.

https://doi.org/10.1371/journal.pgen.1008536.g002

tandem copies of a 13,433 bp sequence from the 5'-flanking region of *ASIP* (Fig 3, S4 Fig, S5 Table).

The Chamois Colored goat and the St. Gallen Booted goat are characterized by a color pattern and an *ASIP* allele termed "badgerface" (A^{b}) [8]. Our pool-seq data revealed a five-fold amplification of 45,680 bp located ~61 kb downstream of the A^{sm} amplification (Fig 3, S4 Fig, S5 Table).

The Peacock goat is a rare Swiss goat breed with a unique and striking coat color pattern that has not been investigated previously. Pool-seq data from Peacock goats indicated a selection signature at the *ASIP* locus. The *ASIP* allele in Peacock goats, which we propose to term "peacock" (A^{Pc}), has a quadruplication of the same ~45 kb region having five copies in the A^{b} allele. It is additionally flanked by triplicated segments of 27,996 bp and 41,807 bp on the left and right side of the quadruplicated sequence (Fig 3, S4 Fig, S5 Table). The central part of the A^{Pc} allele has exactly the same breakpoints as the A^{b} allele suggesting a common origin of A^{b} and A^{Pc}.

The goat genome reference sequence is derived from a San Clemente goat, which has a similar coat color pattern as bezoars. The genome reference therefore supposedly represents the wildtype allele at the *ASIP* locus, termed "bezoar" (A^{bz}) [8]. Bezoars and all other Swiss goat breeds did not show any CNVs at the *ASIP* locus. In the remaining goat breeds the A^{Wt}, A^{sm} and A^{b} alleles were segregating at low frequencies. These breeds are either not specifically

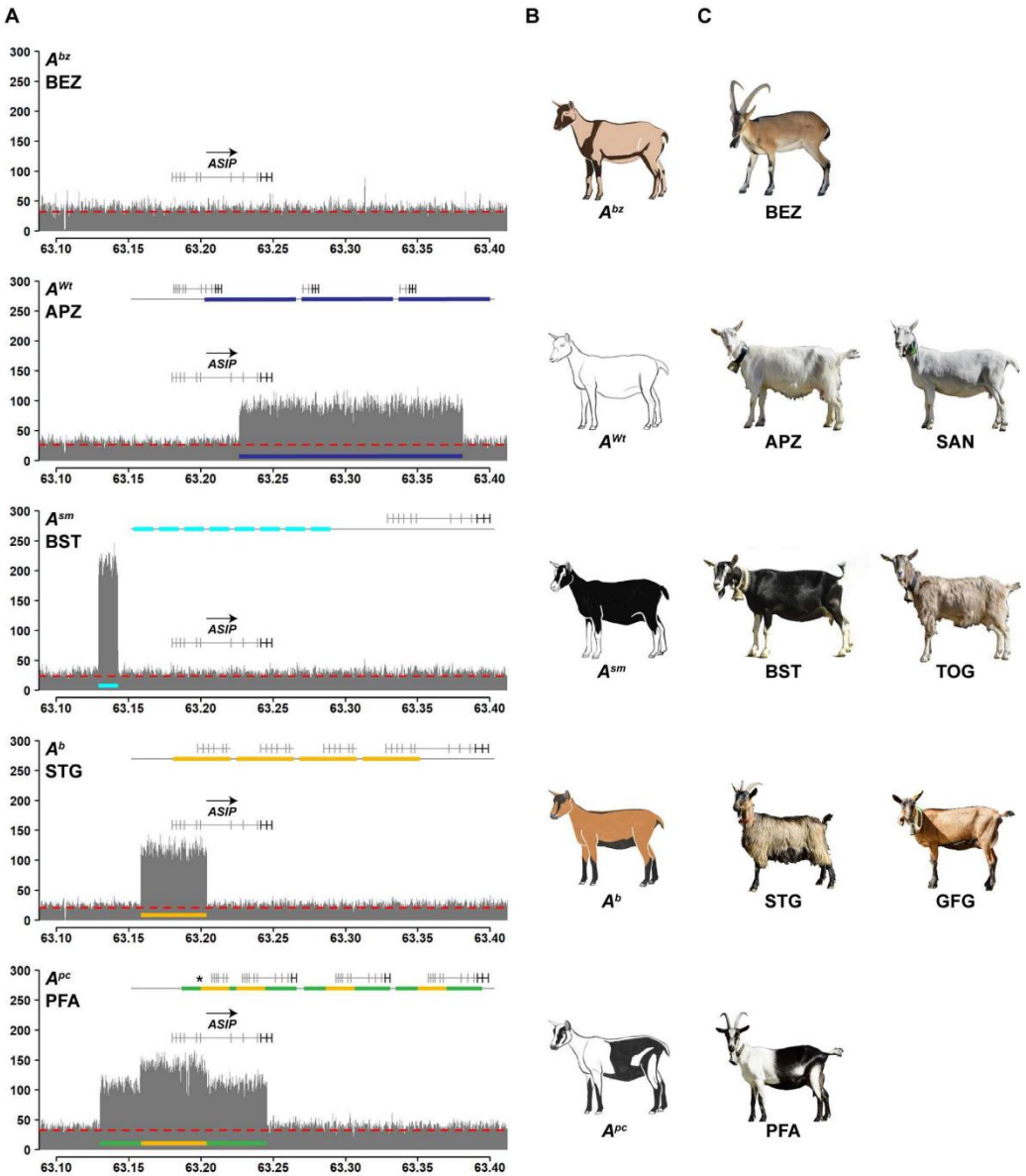

Fig 3. CNVs at the *ASIP* locus. A Coverage plots of the *ASIP* locus in different goat breeds reveal four different CNVs. The bezoar (BEZ) coverage plot shows uniform coverage and is characteristic for the wildtype allele (A^{bz}). Underneath, four different mutant *ASIP* alleles associated with different CNVs are illustrated. The line on top of each plot schematically indicates the most likely configuration of these mutant alleles derived from the available short-read sequence information (S4 Fig). The dashed red line indicates the average coverage across the whole genome of each breed. **B** Schematic drawings and **C** representative photographs illustrating the coat color phenotypes of the studied breeds. The photo of the bezoar was obtained during summer, when the dark stripes at the collar and the belly are much less pronounced than in the winter coat. Note that some of the patterns show an exactly inverse distribution of eumelanin and pheomelanin. For example, goats with the A^{sm} allele have white (pheomelanistic) facial stripes and legs, while goats with the A^{b} or A^{pc} alleles have black (eumelanistic) facial stripes and legs.

https://doi.org/10.1371/journal.pgen.1008536.g003

selected for coat color (e.g. African dwarf goat, Beetal) or the effect of *ASIP* is phenotypically not visible due to epistatic effects of other genes that lead to white spotting phenotypes (e.g. Boer goat, Pak Angora, Barbari).

Quantitative analysis of *ASIP* mRNA expression

The different CNVs in putative regulatory regions of the *ASIP* gene prompted us to hypothesize that quantitative differences in *ASIP* expression may cause the different color patterns. We therefore obtained whole skin samples from five goats carrying different CNV alleles. We isolated RNA from matched pairs of eumelanistic and pheomelanistic skin and performed an RNA-seq experiment to determine the expression level of *ASIP* mRNA expression.

In Grisons Striped goats (A^{sm}), Chamois Colored goats (A^{b}) and Peacock goats (A^{pc}), the eumelanistic skin showed very low *ASIP* mRNA expression. The pheomelanistic skin regions in these three goats had at least 10-fold higher *ASIP* expression than the corresponding eumelanistic samples. The uniformly white (pheomelanistic) Saanen goat (A^{Wt}) had the highest *ASIP* mRNA expression. There was no obvious correlation between the quantitative *ASIP* mRNA expression and the intensity of the pheomelanistic pigmentation. The intensely red colored skin from the Chamois Colored goat had an intermediate *ASIP* mRNA expression compared to the pale white skin from e.g. the Saanen and Peacock goat. Visual inspection of the RNA-seq short-read alignments indicated the utilization of nine different 5'-untranslated exons in nine different transcript isoforms originating from different *ASIP* alleles (Fig 4; S1 File).

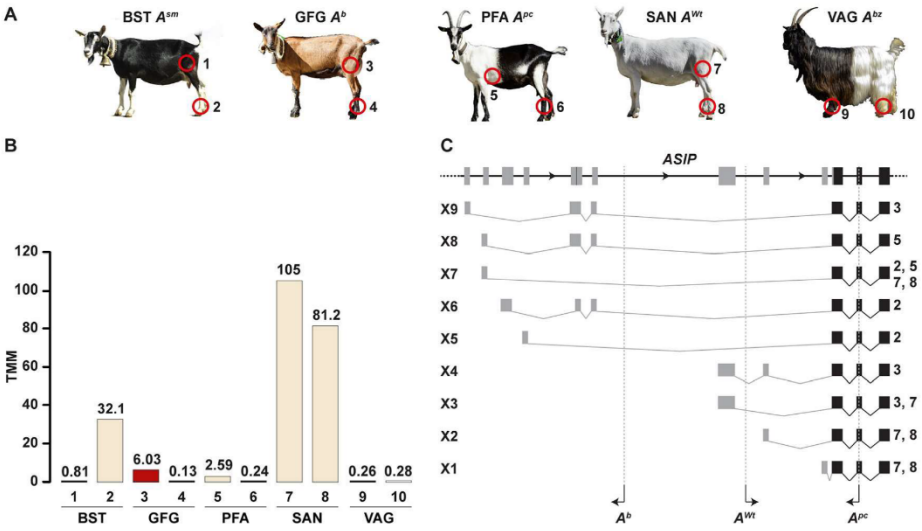

Fig 4. *ASIP* mRNA expression and identified transcripts in skin. A Representative photographs of the five sampled goat breeds. The biopsy sites are numbered and indicated by red circles. **B** Trimmed mean of M (TMM) values of *ASIP* mRNA expression were determined from RNA-seq data for each sample. The colors of the bars correspond to the pigmentation of the skin samples. Please note that the Valais Blackneck goat (VAG) has a black base color that is independent of the *ASIP* gene. This goat has a white spotting phenotype and lacks melanocytes in its caudal half. The low *ASIP* expression in the unpigmented white skin sample of this goat underscores the difference to the pheomelanistic pale white pigmentation in other goats. **C** *ASIP* transcript isoforms in pheomelanistic skin samples from goats with different *ASIP* alleles. Transcript isoforms X1 and X2 correspond to the RefSeq accessions XM_018057735.1 and XM_018057736.1. CNV breakpoints of the A^{b}, A^{Wt}, and A^{pc} alleles are indicated.

https://doi.org/10.1371/journal.pgen.1008536.g004

Discussion

In the present study, we discovered 2,239 loci under selection in 20 diverse goat breeds with various phenotypes and different geographical origins. Our methodology comprised the identification of regions with low heterozygosity from pool-seq data in combination with pairwise F_{ST} to bezoars, the wild ancestor of domesticated goats. The pool-seq approach was validated by repeating the analyses with virtually identical results from individually sequenced goats in five breeds. We have to caution that reduced heterozygosity and high F_{ST} values may not only result from selection, but may also be due to random demographic changes.

The comprehensive catalogue of identified selection signatures can now be used as a starting point to identify causal genetic variants that control a wide variety of breed-defining traits.

We particularly focused on selection signatures harboring known coat color genes and identified two CNVs in the 3'-flanking region of *KIT* in two Pakistani goat breeds, the completely white Pak Angora breed and the white spotted Barbari breed. An association between white coat color and the *KIT* locus has been reported before in Iranian Markhoz goats, which also represents an Angora type goat [37]. The *KIT* gene is flanked by several hundred kilobases of non-coding genomic DNA on either side, which are required for the precise regulation of its temporal and spatial expression. The KIT protein is a receptor tyrosine kinase mediating a survival signal for several different cell types including melanoblasts and melanocytes, but also e.g. hematopoetic stem cells, mast cells, interstitial cells of Cajal and spermatogonia [38,39]. *KIT* is a proto-oncogene and its overexpression may have detrimental consequences such as tumor development [40,41]. Insufficient expression of functional KIT protein in melanoblasts or melanocytes will lead to apoptosis of these cells and results in white spotting phenotypes [15,42–45].

Structural variants at the *KIT* locus cause several other breed defining coat color phenotypes in domestic animals, such as the dominant white and belt phenotypes in pigs [33,46,47], color-sided and lineback in cattle [48,49], and tobiano spotting in horses [50]. Lineback in cattle and tobiano in horses also involve structural variants in the 3'-flanking region of *KIT* [49,50]. All these phenotypes are characterized by striking alterations in pigmentation without any deleterious consequences on the other KIT dependent cell types, which would be expected to result in potentially serious health problems. The comparative data from other species strongly suggest that the newly detected caprine *KIT* CNVs in goats cause the complete lack of skin and hair pigmentation in Pak Angora and the white spotted phenotype in Barbari goats due to altered expression of KIT during fetal development of melanoblasts.

The *ASIP* gene codes for the agouti signaling protein, the competitive inhibitor of melanocortin 1 receptor expressed on melanocytes [20]. Variation in the quantitative amount of *ASIP* mRNA expression from different promoters is the central mechanism regulating the so-called pigment-type switching [19,21,28]. The regulatory elements of the *ASIP* gene are most likely contained in its large 5'-flanking region. Differing from the *KIT* locus, *ASIP* does not contain a very large 3'-flanking region. Spatially and temporally regulated synthesis of eumelanin and pheomelanin enables mammals to express a wide variety of coat color patterns that are essential for e.g. camouflage or mate recognition in many wild species.

ASIP variants cause a wide variety of breed defining coat color phenotypes in domestic animals. *ASIP* loss of function variants, typically in the coding sequence, are responsible for recessive black in e.g. dogs [51], horses [52], and rabbits [53]. Gain of function variants in *ASIP*, such as an ectopic overexpression lead to dominant red phenotypes [54]. There are only very few examples of non-coding regulatory variants at the *ASIP* locus that have been fully characterized at the molecular level. One important example is the mouse black-and tan allele (a^t), which is caused by a ~6 kb retroviral-like insertion in the region of the hair cycle-specific

promoter [25]. In black and tan at mice, hairs are no longer banded and show a uniformly yellow or uniformly black pigmentation. Amplification of the entire *ASIP* gene has previously been shown to cause the white coat color in many sheep breeds [18] and the white or tan allele (A^{Wt}) in Saanen goats [9].

Our study confirmed the previous results and defined the exact breakpoints of the caprine A^{Wt} triplication that actually comprises not only the *ASIP* gene, but also the flanking *AHCY* and *ITCH* genes. Unexpectedly, we did not observe selection signatures at the window harboring the *ASIP* gene in the Saanen and Appenzell breeds. Both of these breeds are strictly selected for uniform white (pheomelanistic) coat color and have a very high frequency of the A^{Wt} allele. We think that the lack of a significant $-ZH_p$ score at *ASIP* in these two breeds is caused by at least three factors. For the calculation of the $-ZH_p$ score, we considered only SNVs with a maximum coverage of 50x in order to suppress artifacts caused by non-specific mapping of highly repetitive sequences. As the *ASIP* locus is triplicated in Saanen and Appenzell goats their pools had >50x average coverage at *ASIP* and almost no SNVs were called in the region. Furthermore, when we inspected the whole genome sequencing data from 24 individual Appenzell goats, we found that only 22 of them were A^{Wt}/A^{Wt} as expected. The remaining two animals had the genotype A^{Wt}/A^{sm}. As A^{Wt} is the most dominant allele in the series, it apparently has still not reached absolute fixation and other *ASIP* alleles are segregating at least in the Appenzell goat breed. Finally, the three copies of the triplication had several sequence differences, which were called as variable positons with a 2:1 ratio of the alleles. This provides a biological explanation why the selection signature might be weaker than expected.

We observed significant selection signatures at the *ASIP* gene in five Swiss goat breeds. In these breeds, we identified three additional non-coding CNVs in the 5'-region of *ASIP* that are likely to cause the Swiss markings (A^{sm}), badgerface (A^{b}) and peacock (A^{Pc}) alleles in goats. We have to caution that the peacock phenotype has not been reported before and may be influenced by additional genes other than *ASIP*.

The corresponding coat color phenotypes are very interesting as they have characteristic patterns of eumelanistic and pheomelanistic pigmentation. Domestic goats with these alleles therefore represent a valuable resource for dissecting the precise function of individual regulatory elements in future studies. The A^{sm} and A^{b} alleles result in almost exactly inverted distributions of eumelanin and pheomelanin. Our RNA-seq data confirm that the different pigmentation patterns are caused by different levels of *ASIP* mRNA transcription at different body locations. These data also revealed that the different caprine *ASIP* gene alleles give rise to a higher number of non-coding 5'-exons compared to other mammals [25,26]. Additional data, such as Cage-seq and full-length Iso-seq data will be required for a comprehensive annotation of all possible transcription start sites and splice isoforms in goats. Such data are expected to become available soon with the advances of the FAANG project [55].

In conclusion, we identified 2,239 selection signatures in 20 diverse goat breeds with various coat color phenotypes. These selection signatures revealed six different functionally relevant CNVs underlying breed-defining coat color phenotypes in goats. The results should help to advance our mechanistic understanding of temporal and spatial regulation of transcription.

Materials and methods

Ethics statement

All animal experiments were performed according the local regulations. All animals in this study were examined with the consent of their owners. Sample collection was approved by the "Cantonal Committee For Animal Experiments" (Canton of Bern; permit 75/16).

Animals

For this study, 244 female animals of 20 phenotypical diverse goat breeds and their wild ancestor, the bezoar, were sampled (S1 Table). Ten of the analyzed goat breeds originate from Switzerland, eight from Pakistan and two from Africa. Swiss and African breeds were sampled in Switzerland. Pakistani breeds were sampled in Pakistan. bezoar samples were from zoo animals. For the Swiss goat breeds, we selected representative animals of the breeds and excluded any first-degree relatives. For the other goat breeds, we did not have full pedigree information and used convenience samples. Genomic DNA was isolated from EDTA blood samples.

Whole genome sequencing of pools (pool-seq)

Breed pools were prepared by pooling equimolar amounts of 12 animals per breed (ANG: 10, BEZ: 8 and STG: 10). Illumina TruSeq PCR-free genomic DNA libraries with an insert size of 350 bp were prepared. Each breed pool was sequenced on one lane of an Illumina HiSeq 3000 instrument and on average 300 million 2x150 bp paired-end reads per breed pool were collected (S1 Table).

Mapping and variant calling

Adapter sequences, reads with too many Ns and low quality bases were trimmed or ultimately discarded, if the remaining read length was < 50 bp with fastq-mcf version 1.1.2 (settings: -l 50 -S -q 20). The cleaned reads were mapped to the goat reference genome ARS1 [56] with Burrows-Wheeler Aligner (BWA-MEM) algorithm version 0.7.13 [57] using the "-M" flag to mark shorter alignments as secondary. The resulting mapping files in SAM format were converted into BAM format and coordinate sorted using SAMtools version 1.3 [58]. A local indel realignment was performed using the Genome Analysis Toolkit version 3.7 [59] with default settings. Duplicated reads were marked, using Picard Tools version 2.2.1 (http://broadinstitute.github.io/picard) with default settings for patterned flow cell models. Single nucleotide variants were called using (i) Genome Analysis Toolkit UnifiedGenotyper version 3.7 [59] with the settings: -glm SNP, -stand_call_conf 20, -out_mode EMIT_VARIANTS_ONLY and –ploidy 16/20/24 and (ii) SAMtools mpileup [60] with the settings -q 15, -Q 20, -C 50 and -B. The variants resulting from UnifiedGenotyper were filtered for high quality variants with GATK's VariantFiltration tool using the generic hard-filtering recommendations available from https://gatkforums.broadinstitute.org/gatk/discussion/6925/understanding-and-adapting-the-generic-hard-filtering-recommendations, while the mpileup files were streamed to the PoPoolation2 version 1.201 pipeline [61]. We used the scripts mpileup2sync.jar with settings --fastq-type sanger and --min-qual 20 and snp-frequency-diff.pl with the settings --min-coverage 15, --max-coverage 50 and --min-count 3. Both pipelines yielded similar numbers of SNV (S2 Table).

Sweep analysis of pool-seq data

A screen for selective sweeps was performed using the SNV file produced for each breed pool individually by the mpileup PoPoolation2 pipeline. At each identified SNV position in the files, we took the numbers of major (n_{MAJ}) and minor (n_{MIN}) allele counts observed in each breed and calculated pooled heterozygosity (H_p) [32] with an in-house written script. The script applies $H_p = 2\Sigma n_{MAJ}\Sigma n_{MIN}/(\Sigma n_{MAJ}+\Sigma n_{MIN})^2$ in a sliding 50% overlapping window approach. We evaluated the results with different window sizes (25 to 300 kb) and decided on 150 kb as the most appropriate size [33, 34]. The obtained H_p values for all 34,382 overlapping 150 kb windows across the whole genome were Z transformed, performing $ZH_p = (H_p\text{-}\mu H_p/$

σH_p). Windows with a $-ZH_p \geq 4$ were retained as selective windows and adjacent or overlapping selective windows were merged into selection signatures, individually per breed. We annotated the identified selection signatures along with NCBI's *Capra hircus* Annotation Release 102 (S3 Table). In addition, to the H_p calculation, we calculated weighted population F_{ST} values for each SNV in a 150 kb sliding, 50% overlapping window approach. We applied the F_{ST}-sliding.pl script of the Popoolation2 pipeline. The script F_{ST}-sliding.pl was run with the settings --min-count 2 --min-coverage 4 --max-coverage 50 --window-size 150000 --step-size 75000 --suppress-noninformative and --pool-size 10:12:12:12:8:12:12:12:12:12:12:12:12:12:12:12:12:10:12:12:12:12. It used the previously obtained sync file of all pools combined as input and calculated weighted population pairwise F_{st} values using the standard equation as shown in Hartl and Clark [62]. This resulted in 210 pairwise comparisons, from which we selected the comparisons between the 20 domesticated goat breeds with the bezoar. The obtained F_{ST} values were Z transformed, performing $ZF_{ST} = (F_{ST}-\mu F_{ST}/\sigma F_{ST})$.

Whole genome re-sequencing of individual goats and variant calling

In addition to the pool-seq experiment, we selected 120 goats from five Swiss breeds for individual whole genome re-sequencing. The 24 animals per breed included the 12 goats represented in the breed pools (S1 Table). Illumina TruSeq PCR-free DNA libraries with an insert size of 350 bp were prepared and sequenced on an Illumina NovaSeq 6000 instrument, yielding on average 240 million 2x150 bp paired-end reads per goat (S1 Table). Clean reads were produced by running fastp, version 0.12.5 [63], an ultra-fast all-in-one FASTQ preprocessor capable of trimming polyG tails, a known issue of NovaSeq reads. The cleaned reads were mapped to the ARS1 goat reference genome [56] with Burrows-Wheeler Aligner (BWA-MEM) algorithm using the "-M" flag to mark shorter alignments as secondary. The resulting SAM files were converted into BAM files and coordinate sorted using SAMtools. Duplicated reads were marked, using Picard Tools (http://broadinstitute.github.io/picard) with default settings for patterned flow cell models. The marked BAM files were streamed to GATK's Base-Recalibrator tool, supported with known SNV provided by the VarGoats consortium (http://www.goatgenome.org/vargoats.html). Subsequently, GATK's HaplotypeCaller with the settings --emitRefConfidence GVCF and -stand_call_conf 30 was used to call genome-wide variants [59]. The variant files were merged and GATK's GenotypeGVCFs was used to call variants in the 120 goats combined. As a next step, the called variants were filtered for high quality variants with GATK's VariantFiltration tool (version 3.8) using the generic hard-filtering recommendations available from https://gatkforums.broadinstitute.org/gatk/discussion/6925/understanding-and-adapting-the-generic-hard-filtering-recommendations. SnpEff [64] and NCBI's *Capra hircus* Annotation Release 102 was used to annotate the variants.

Sweep analysis of individual goats

To calculate H_p scores of the individual goats, we selected biallelic, passed SNPs per breed using GATK's SelectVariants tool (version 3.8), applying --restrictAllelesTo BIALLELIC --selectTypeToInclude SNP --sample_expressions '(APZ/BST/PFA/STG/VAG)' --maxNOCALLnumber 0 --excludeFiltered --excludeNonVariants. This yielded on average 14.9 million SNVs per combined Swiss goat breed, comprising each 24 individually sequenced animals (S2 Table). As a next step, the VCF files containing only biallelic, passed SNPs were transformed into table format using GATK's VariantsToTable tool. This table contained only information regarding SNP position, reference allele and genotype of the 24 animals. With an in-house written Python script, we converted the table produced with GATK's VariantsToTable to

major and minor alleles and counted the number of observations. This output was then used for H_p calculation as described in sweep analysis of pool-seq data.

Manhattan plots

The $-ZH_p$ values were plotted using the function manhattan of the qqman package [65] with R [66]. Each data point represents a 150 kb window. A red horizontal line was drawn representing the chosen significance threshold of $-ZH_p \geq 4$ (corresponding to 0.8% of all windows).

CNV analyses

Coverage plots for regions of interest were created by calculating the coverage of each base in a defined region of interest using Samtools depth -b. Additionally coverage stats across the whole genomes, including the average coverage were calculated using goleft covstats (https://github.com/brentp/goleft). Taken both results together, we plotted the coverage using R plot type h version 3.4.1 and indicated the average coverage line. Potential CNVs were also visually evaluated by inspection of the short-read alignemnts (bam-files) in the Integrative Genome Viewer (IGV) [67].

Skin biopsies and total RNA extraction

Skin biopsies were taken from five slaughtered animals of different goat breeds (SAN, BST, GFG, PFA and VAG). Two 6 mm punch biopsies were taken from differentially pigmented body areas of each animal (S6 Table). The biopsies were immediately put in RNA*later* (Qiagen) for at least 24 h and then frozen at −20˚C. Prior to RNA extraction, the skin biopsies were homogenized mechanically with the TissueLyser II device from Qiagen. Total RNA was extracted from the homogenized tissue using the RNeasy Fibrous Tissue Mini Kit (Qiagen) according to the manufacturer's instructions. RNA quality was assessed with a FragmentAnalyzer (Advanced Analytical) and the concentration was measured using a Qubit Fluorometer (ThermoFisher Scientific).

Whole transcriptome sequencing (RNA-seq)

From each sample, 1 μg of high quality total RNA (RIN >9) was used for library preparation with the Illumina TruSeq Stranded mRNA kit. The 10 libraries were pooled and sequenced on an S1 flow cell with 2x50 bp paired-end sequencing using an Illumina NovaSeq 6000 instrument. On average, 31.5 million paired-end reads per sample were collected (S6 Table). All reads that passed quality control were mapped to the ARS1 goat reference genome assembly using STAR aligner (version 2.6.0c) [68]. The read abundance was calculated using HTseq (version 0.9.1) [69] and a gff3 file obtained from NCBI's *Capra hircus* Annotation Release 102. We used the EdgeR package [70] to read the HTseq count data and calculated the log fold changes using the exactTest function where the biological co-efficient of variation (BCV) was set to 0.1. Trimmed mean of M (TMM) values of *ASIP* mRNA expression were determined for each sample [71].

Supporting information

S1 Fig. Distributions of H_p, $-ZH_p$, F_{ST} and $-ZF_{ST}$.
(PDF)

S2 Fig. $-ZH_p$ scores Manhattan plots of pooled sequencing and single sequencing.
(PDF)

S3 Fig. Manhattan plots of −ZHp scores and ZF_{ST} scores.
(PDF)

S4 Fig. Details of the CNV alleles.
(PDF)

S1 File. FASTA sequences of nine different caprine *ASIP* transcripts.
(TXT)

S1 Table. Read statistics and accessions of pool-seq and individual WGS data.
(XLSX)

S2 Table. Pool-seq and individual WGS SNV statistics.
(XLSX)

S3 Table. H_p selection signatures per breed pool.
(XLSX)

S4 Table. F_{ST} selection signatures per breed pool.
(XLSX)

S5 Table. Details of CNV alleles.
(XLSX)

S6 Table. Descriptive statistics and accessions of RNA-seq datasets.
(XLSX)

Acknowledgments

The authors are grateful to all goat owners and breeding organizations who donated samples and shared pedigree data and phenotype information of their animals. We thank Eva Andrist, Nathalie Besuchet Schmutz, Muriel Fragnière, and Sabrina Schenk for expert technical assistance, the Next Generation Sequencing Platform of the University of Bern for performing the high-throughput sequencing experiments, and the Interfaculty Bioinformatics Unit of the University of Bern for providing high performance computing infrastructure. Furthermore, we thank Christian Gazzarin for the photos and Sarah Stangl for the graphical illustrations of the different Swiss goat breeds.

Author Contributions

Conceptualization: Tosso Leeb.

Data curation: Vidhya Jagannathan.

Funding acquisition: Tosso Leeb.

Investigation: Jan Henkel, Rashid Saif, Vidhya Jagannathan, Corinne Schmocker, Flurina Zeindler, Cord Drögemüller, Christine Flury.

Methodology: Vidhya Jagannathan.

Resources: Rashid Saif, Erika Bangerter, Ursula Herren, Dimitris Posantzis, Zafer Bulut, Philippe Ammann, Cord Drögemüller.

Supervision: Vidhya Jagannathan, Cord Drögemüller, Christine Flury, Tosso Leeb.

Visualization: Jan Henkel.

Writing – original draft: Jan Henkel, Tosso Leeb.

Writing – review & editing: Jan Henkel, Rashid Saif, Vidhya Jagannathan, Corinne Schmocker, Flurina Zeindler, Erika Bangerter, Ursula Herren, Dimitris Posantzis, Zafer Bulut, Philippe Ammann, Cord Drögemüller, Christine Flury, Tosso Leeb.

References

1. Zeder MA, Hesse B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. Science. 2000; 287: 2254–2257. https://doi.org/10.1126/science.287.5461.2254 PMID: 10731145
2. Naderi S, Rezaei H-R, Pompanon F, Blum MG, Negrini R, Naghash H-R, et al. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proc Natl Acad Sci USA. 2008; 105: 17659–17664. https://doi.org/10.1073/pnas.0804782105 PMID: 19004765
3. Colli L, Milanesi M, Talenti A, Bertolini F, Chen M, Crisa A, et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. Genet Sel Evol. 2018; 50: 58. https://doi.org/10.1186/s12711-018-0422-x PMID: 30449284
4. FAO. The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture, edited by B.D. Scherf & D. Pilling. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome 2015. (http://www.fao.org/3/a-i4787e/index.html)
5. Stella A, Nicolazzi EL, Van Tassell CP, Rothschild MF, Colli L, Rosen BD, et al. AdaptMap: exploring goat diversity and adaptation. Genet Sel Evol. 2018; 50: 61. https://doi.org/10.1186/s12711-018-0427-5 PMID: 30453882
6. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. Nat Comm. 2018; 9: 813.
7. Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, et al. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. Anim Genet. 2016; 47: 727–739. https://doi.org/10.1111/age.12476 PMID: 27436146
8. Adalsteinsson S, Sponenberg DP, Alexieva S, Russel AJ. Inheritance of goat coat colors. J Hered. 1994; 85: 267–272. https://doi.org/10.1093/oxfordjournals.jhered.a111454 PMID: 7930499
9. Fontanesi L, Beretti F, Riggio V, Gomez Gonzalez E, Dall'Olio S, Davoli R, et al. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. Cytogenet Genome Res. 2009; 126:333–347. https://doi.org/10.1159/000268089 PMID: 20016133
10. Becker D, Otto M, Ammann P, Keller I, Drögemüller C, Leeb T. The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the *TYRP1* locus on chromosome 8. Anim Genet. 2015; 46:50–54. https://doi.org/10.1111/age.12240 PMID: 25392961
11. Dietrich J, Menzi F, Ammann P, Drögemüller C, Leeb T. A breeding experiment confirms the dominant mode of inheritance of the brown coat colour associated with the ^{496}Asp *TYRP1* allele in goats. Anim Genet. 2015; 46: 587–588. https://doi.org/10.1111/age.12320 PMID: 26153465
12. Menzi F, Keller I, Reber I, Beck J, Brenig B, Schutz E, et al. Genomic amplification of the caprine *EDNRA* locus might lead to a dose dependent loss of pigmentation. Sci Rep. 2016; 6: 28438. https://doi.org/10.1038/srep28438 PMID: 27329507
13. Jackson IJ. Homologous pigmentation mutations in human, mouse and other model organisms. Hum Mol Genet. 1997; 6: 1613–1624. https://doi.org/10.1093/hmg/6.10.1613 PMID: 9300652
14. Thomas AJ, Erickson CA. The making of a melanocyte: the specification of melanoblasts from the neural crest. Pigment Cell Melanoma Res. 2008; 21: 598–610. https://doi.org/10.1111/j.1755-148X.2008.00506.x PMID: 19067969
15. Haase B, Brooks SA, Schlumbaum A, Azor PJ, Bailey E, Alaeddine F, et al. Allelic heterogeneity at the equine *KIT* locus in dominant white (*W*) horses. PLoS Genet. 2007; 3: e195. https://doi.org/10.1371/journal.pgen.0030195 PMID: 17997609
16. Greenhill ER, Rocco A, Vibert L, Nikaido M, Kelsh RN. An iterative genetic and dynamical modelling approach identifies novel features of the gene regulatory network underlying melanocyte development. PLoS Genet. 2011; 7: e1002265. https://doi.org/10.1371/journal.pgen.1002265 PMID: 21909283
17. Hauswirth R, Haase B, Blatter M, Brooks SA, Burger D, Drögemüller C, et al. Mutations in *MITF* and *PAX3* cause "splashed white" and other white spotting phenotypes in horses. PLoS Genet. 2012; 8: e1002653. https://doi.org/10.1371/journal.pgen.1002653 PMID: 22511888

18. Norris BJ, Whan VA. A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep. Genome research. 2008; 18(8):1282–93. https://doi.org/10.1101/gr.072090.107 PMID: 18493018

19. Barsh GS. The genetics of pigmentation: from fancy genes to complex traits. Trends Genet. 1996; 12: 299–305. https://doi.org/10.1016/0168-9525(96)10031-7 PMID: 8783939

20. Lu D, Willard D, Patel IR, Kadwell S, Overton L, Kost T, et al. Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. Nature. 1994; 371: 799–802. https://doi.org/10.1038/371799a0 PMID: 7935841

21. Millar SE, Miller MW, Stevens ME, Barsh GS. Expression and transgenic studies of the mouse agouti gene provide insight into the mechanisms by which mammalian coat color patterns are generated. Development. 1995; 121: 3223–3232. PMID: 7588057

22. Le Pape E, Wakamatsu K, Ito S, Wolber R, Hearing VJ. Regulation of eumelanin/pheomelanin synthesis and visible pigmentation in melanocytes by ligands of the melanocortin 1 receptor. Pigment Cell Melanoma Res. 2008; 21: 477–486. https://doi.org/10.1111/j.1755-148X.2008.00479.x PMID: 18627531

23. Cieslak M, Reissmann M, Hofreiter M, Ludwig A. Colours of domestication. Biol Rev. 2011; 86: 885–899. https://doi.org/10.1111/j.1469-185X.2011.00177.x PMID: 21443614

24. Graham A, Wakamatsu K, Hunt G, Ito S, Thody AJ. Agouti protein inhibits the production of eumelanin and pheomelanin in the presence and absence of alpha-melanocyte stimulating hormone. Pigment Cell Res. 1997; 10: 298–303. https://doi.org/10.1111/j.1600-0749.1997.tb00689.x PMID: 9359625

25. Bultman SJ, Michaud EJ, Woychik RP. Molecular characterization of the mouse agouti locus. Cell. 1992; 71: 1195–1204. https://doi.org/10.1016/s0092-8674(05)80067-4 PMID: 1473152

26. Vrieling H, Duhl DM, Millar SE, Miller KA, Barsh GS. Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. Proc Natl Acad Sci USA. 1994; 91: 5667–5671. https://doi.org/10.1073/pnas.91.12.5667 PMID: 8202545

27. Drögemüller C, Giese A, Martins-Wess F, Wiedemann S, Andersson L, Brenig B, et al. The mutation causing the black-and-tan pigmentation phenotype of Mangalitza pigs maps to the porcine *ASIP* locus but does not affect its coding sequence. Mamm Genome. 2006; 17: 58–66. https://doi.org/10.1007/s00335-005-0104-1 PMID: 16416091

28. Kaelin CB, Barsh GS. Genetics of pigmentation in dogs and cats. Ann Rev Anim Biosci. 2013; 1: 125–156.

29. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974; 23: 23–35. PMID: 4407212

30. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics. 2005; 169: 2335–2352. https://doi.org/10.1534/genetics.104.036947 PMID: 15716498

31. Pennings PS, Hermisson J. Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol 2006; 23: 1076–1084. https://doi.org/10.1093/molbev/msj117 PMID: 16520336

32. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010; 464: 587–591. https://doi.org/10.1038/nature08832 PMID: 20220755

33. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci USA. 2012; 109: 19529–19536. https://doi.org/10.1073/pnas.1217149109 PMID: 23151514

34. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alfoldi J, Martinez Barrio A, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 2014; 345: 1074–1079. https://doi.org/10.1126/science.1253714 PMID: 25170157

35. Martinez Barrio A, Lamichhaney S, Fan G, Rafati N, Pettersson M, Zhang H, et al. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. eLife. 2016; 5: e12081. https://doi.org/10.7554/eLife.12081 PMID: 27138043

36. Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep. 2016; 6: 38932. https://doi.org/10.1038/srep38932 PMID: 27941843

37. Nazari-Ghadikolaei A, Mehrabani-Yeganeh H, Miarei-Aashtiani SR, Staiger EA, Rashidi A, Huson HJ. Genome-wide association studies identify candidate genes for coat color and mohair traits in the Iranian Markhoz goat. Front Genet. 2018; 9: 105. https://doi.org/10.3389/fgene.2018.00105 PMID: 29670642

38. Chabot B, Stephenson DA, Chapman VM, Besmer P, Bernstein A. The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. Nature. 1988; 335: 88–89. https://doi.org/10.1038/335088a0 PMID: 2457811

39. Roskoski R Jr. Signaling by Kit protein-tyrosine kinase–the stem cell factor receptor. Biochem Biophys Res Commun. 2005; 337: 1–13. https://doi.org/10.1016/j.bbrc.2005.08.055 PMID: 16129412

40. Furitsu T, Tsujimura T, Tono T, Ikeda H, Kitayama H, Koshimizu U, et al. Identification of mutations in the coding sequence of the proto-oncogene c-kit in a human mast cell leukemia cell line causing ligand-independent activation of c-kit product. J Clin Invest. 1993; 92: 1736–1744. https://doi.org/10.1172/JCI116761 PMID: 7691885

41. Nagata H, Worobec AS, Semere T, Metcalfe DD. Elevated expression of the proto-oncogene c-kit in patients with mastocytosis. Leukemia. 1998; 12: 175–181. https://doi.org/10.1038/sj.leu.2400906 PMID: 9519779

42. Geissler EN, Ryan MA, Housman DE. The dominant-white spotting (*W*) locus of the mouse encodes the c-kit proto-oncogene. Cell. 1988; 55: 185–192. https://doi.org/10.1016/0092-8674(88)90020-7 PMID: 2458842

43. Giebel LB, Spritz RA. Mutation of the *KIT* (mast/stem cell growth factor receptor) protooncogene in human piebaldism. Proc Nat Acad Sci USA. 1991; 88: 8696–8699. https://doi.org/10.1073/pnas.88.19.8696 PMID: 1717985

44. Fleischman RA, Saltman DL, Stastny V, Zneimer S. Deletion of the c-kit protooncogene in the human developmental defect piebald trait. Proc Natl Acad Sci USA. 1991; 88: 10885–10889. https://doi.org/10.1073/pnas.88.23.10885 PMID: 1720553

45. Haase B, Brooks SA, Tozaki T, Burger D, Poncet PA, Rieder S, et al. Seven novel *KIT* mutations in horses with white coat colour phenotypes. Anim Genet. 2009; 40: 623–629. https://doi.org/10.1111/j.1365-2052.2009.01893.x PMID: 19456317

46. Marklund S, Kijas J, Rodriguez-Martinez H, Rönnstrand L, Funa K, Moller M, et al. Molecular basis for the dominant white phenotype in the domestic pig. Genome Res. 1998; 8: 826–833. https://doi.org/10.1101/gr.8.8.826 PMID: 9724328

47. Giuffra E, Evans G, Törnsten A, Wales R, Day A, Looft H, Plastow G, Andersson L. The Belt mutation in pigs is an allele at the Dominant white (*I/KIT*) locus. Mamm Genome. 1999; 10: 1132–1136. https://doi.org/10.1007/s003359901178 PMID: 10594235

48. Durkin K, Coppieters W, Drögemüller C, Ahariz N, Cambisano N, Druet T, et al. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature. 2012; 482: 81–84. https://doi.org/10.1038/nature10757 PMID: 22297974

49. Küttel L, Letko A, Häfliger IM, Signer-Hasler H, Joller S, Hirsbrunner G, et al. A complex structural variant at the *KIT* locus in cattle with the Pinzgauer spotting pattern. Anim Genet. 2019; https://doi.org/10.1111/age.12821 PMID: 31294880

50. Brooks SA, Lear TL, Adelson DL, Bailey E. A chromosome inversion near the *KIT* gene and the Tobiano spotting pattern in horses. Cytogenetic and genome research. 2007; 119: 225–230. https://doi.org/10.1159/000112065 PMID: 18253033

51. Kerns JA, Newton J, Berryere TG, Rubin EM, Cheng JF, Schmutz SM, et al. Characterization of the dog Agouti gene and a nonagoutimutation in German Shepherd Dogs. Mamm Genome. 2004; 15: 798–808. https://doi.org/10.1007/s00335-004-2377-1 PMID: 15520882

52. Rieder S, Taourit S, Mariat D, Langlois B, Guerin G. Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (*Equus caballus*). Mamm Genome. 2001; 12: 450–455. https://doi.org/10.1007/s003350020017 PMID: 11353392

53. Fontanesi L, Forestier L, Allain D, Scotti E, Beretti F, Deretz-Picoulet S, et al. Characterization of the rabbit agouti signaling protein (ASIP) gene: transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. Genomics. 2010; 95: 166–175. https://doi.org/10.1016/j.ygeno.2009.11.003 PMID: 20004240

54. Berryere TG, Kerns JA, Barsh GS, Schmutz SM. Association of an Agouti allele with fawn or sable coat color in domestic dogs. Mamm Genome. 2005; 16: 262–272. https://doi.org/10.1007/s00335-004-2445-6 PMID: 15965787

55. Giuffra E, Tuggle CK; FAANG Consortium. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. Annu Rev Anim Biosci. 2019; 7: 65–88. https://doi.org/10.1146/annurev-animal-020518-114913 PMID: 30427726

56. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nature Genet. 2017; 49: 643–650. https://doi.org/10.1038/ng.3802 PMID: 28263316

57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

59. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20: 1297–1303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199

60. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

61. Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011; 27: 3435–3456. https://doi.org/10.1093/bioinformatics/btr589 PMID: 22025480

62. Hartl DL, Clark AG. Principles of population genetics: Sinauer associates Sunderland, MA; 1997.

63. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34: i884–i90. https://doi.org/10.1093/bioinformatics/bty560 PMID: 30423086

64. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012; 6: 80–92. https://doi.org/10.4161/fly.19695 PMID: 22728672

65. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. 2014:005165.

66. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2018. https://www.R-project.org/.

67. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinf. 2013; 14: 178–192.

68. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

69. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31: 166–169. https://doi.org/10.1093/bioinformatics/btu638 PMID: 25260700

70. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

71. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010; 11: R25. https://doi.org/10.1186/gb-2010-11-3-r25 PMID: 20196867

4.2 Introgression of *ASIP* and *TYRP1* alleles explains coat color variation in Valais goats

Journal:	Journal of Heredity
Manuscript status:	accepted
Contributions:	genetic analyses, illustrations, original draft, review and editing of manuscript

Displayed version: Submitted
DOI:

This article has been accepted for publication in Journal of Heredity Published by Oxford University Press.

Brief Communication

Introgression of *ASIP* and *TYRP1* alleles explains coat color variation in Valais goats

Jan Henkel[1,2], Alexandra Dubacher[1], Erika Bangerter[3], Ursula Herren[3], Philippe Ammann[4], Cord Drögemüller[1,2], Christine Flury[5], Tosso Leeb[1,2*]

[1] Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland

[2] DermFocus, University of Bern, 3001 Bern, Switzerland

[3] Swiss Goat Breeding Association, 3052 Zollikofen, Switzerland

[4] ProSpecieRara, 4052 Basel, Switzerland

[5] School of Agricultural, Forest and Food Sciences, Bern University of Applied Sciences, 3052 Zollikofen, Switzerland

*Address for correspondence:

Tosso Leeb
Institute of Genetics
University of Bern
Bremgartenstrasse 109a
3001 Bern
Switzerland

Phone +41 31 631 2326
Fax +41 31 631 2640
E-mail Tosso.Leeb@vetsuisse.unibe.ch

Running title: Coat color allele introgression in Valais goats

Abstract

The Valais Blackneck goat is a Swiss goat breed with a characteristic coat color phenotype. Before the revision of the breed standard in 1938, four different color varieties of Valais goats were known. Besides Blackneck animals resembling the modern breed standard, the brown and white Copperneck goat, the white Capra Sempione and the greyish Grüenochte comprised the historic Valais goats. The brown pigmentation of Copperneck goats had previously been traced back to an introgression of a mutant *TYRP1* allele from Toggenburg goats. In the present study, we identified additional introgression events of distinct *ASIP* alleles causing the remaining two rare coat color patterns within the Valais Blackneck goat breed. We identified the introgression of the A^{Wt} allele from Appenzell or Saanen goats in white Capra Sempione goats. Similarly, introgression of the A^{Pc} allele from Peacock goats resulted in the greyish Grüenochte phenotype. These results demonstrate past hybridization events between breeds that are separated today. A perfect genotype-phenotype association in 393 Valais goats supported the causality of the genotyped variants for the different coat color phenotypes. Our study gives insights into the introgression of functionally relevant copy number variant (CNV) alleles controlling pigmentation between goat breeds with strikingly different coat color patterns.

Keywords: *Capra hircus*, gene flow, pattern, copy number variant, CNV, agouti

Introduction

Artificial selection of domesticated goats resulted in a variety of diverse goat breeds selected not only for milk, meat or fibers, but also for unique coat color phenotypes (FAO 2015; Burren *et al.* 2016). Due to their striking appearance, these goats were of special value to their owners and selected during breed formation. In many cases, coat color was the primary characteristic of new breeds, which subsequently developed into closed populations (Andersson 2013; Andersson 2016). Coat color variation was intensively studied in goats and several causative genetic variants for specific base colors or color patterns were reported (Adalsteinsson *et al.* 1994, Fontanesi *et al.* 2009; Becker *et al.* 2015; Dietrich *et al.* 2015; Menzi *et al.* 2016; Henkel *et al.* 2019).

Switzerland has ten local goat breeds showing diverse coat color phenotypes (Glowatzki-Mullis *et al.* 2008; Burren *et al.* 2016; Henkel *et al.* 2019). The Valais goat population originally comprised four color variants (Figure 1). The most common of these is the Valais Blackneck goat characterized by solid black pigmentation of the cranial half of the animal separated by a sharp vertical boundary of the depigmented white caudal half. The Blackneck phenotype became the only allowed phenotype of the modern breed standard in 1938. The other coat color variants of the Valais goats are the Copperneck goat, characterized by a brown and white pattern, the completely white Capra Sempione or Simplonziege and the greyish Grüenochte. The three rare coat color variants were maintained by a few breeders and are now regaining increasing popularity. ProSpecieRara, a non-governmental organization dedicated to the conservation of rare local breeds established a separate herdbook comprising the three rare coat color variants of Valais goats in the year 2013 (www.prospecierara.ch).

We previously reported that the brown coat color of the Copperneck goat is due to the introgression of the dominant *TYRP1*^{496Asp} allele from Toggenburg goats (Becker *et al.* 2015; Dietrich *et al.* 2015). The goal of this study was the clarification of the genetic variation causing the coat color phenotypes in the white Capra Sempione and the greyish Grüenochte.

Figure 1. Coat color varieties of Valais goats.

Materials and Methods

Ethics statement

All animal experiments were performed according the local regulations. All animals in this study were examined with the consent of their owners. Sample collection was approved by the "Cantonal Committee For Animal Experiments" (Canton of Bern; permit 71/19).

Breed definition and *ASIP* allele nomenclature

This study investigated four distinct coat color varieties of Valais goats, termed Blackneck, Copperneck, Capra Sempione and Grüenochte. Valais Blackneck goats form a major Swiss goat breed whose breed standard was revised in 1938 to include only the Blackneck coat color variety. In addition to the Valais Blackneck goat herdbook, another smaller herdbook of Valais goats

comprising the three rare coat color varieties was started in 2013. *ASIP* allele nomenclature was used as described previously (Adalsteinsson *et al.* 1994; Henkel *et al.* 2019).

Animals and sequencing

For this study, a total of 393 goat samples were used (Supplementary Table S1). Four representative, unrelated, female Valais goats of the different coat color patterns were selected for whole genome sequencing. Illumina TruSeq PCR-free DNA libraries with an insert size of 350 bp were prepared and sequenced on an Illumina HiSeq 2000 or NovaSeq 6000 instrument (Supplementary Table S2). On average, 200 million paired-end reads per goat were collected. The sequence data of the Valais Blackneck goat were already reported in an earlier publication of our group (Henkel *et al.* 2019).

Mapping and Coverage plots

Mapping of the raw sequence data to the ARS1 reference genome assembly (Bickhart *et al.* 2017) was performed as described in Henkel *et al.* 2019. We visually inspected the short-read alignments in the Integrative Genome Viewer (IGV) to assess large structural variants (Thorvaldsdottir *et al.* 2013). We calculated the coverage of each base in a defined region of interest using samtools depth -b (Li H *et al.* 2009). Additionally coverage stats across the whole genome, including the average coverage were calculated using goleft covstats (https://github.com/brentp/goleft). We plotted the coverage using R plot type h version 3.4.1 (R Core Team 2018).

Genotyping

We designed primer pairs spanning the junction of the tandem copies of two CNVs in the *ASIP* gene to detect the presence of the mutant A^{Wt} and A^{Pc} alleles (Supplementary Table S3). PCR was performed using the primers specific for A^{Wt} or A^{Pc} and ATG360 polymerase (ThermoFischer). The PCR products were analyzed using a 5200 Fragment Analyzer System (Agilent). These assays detected the presence of the A^{Wt} or A^{Pc} alleles, but could not discriminate between homozygous

mutant or heterozygous animals. We genotyped the *TYRP1* variant NM_001285727.1:c.1487G>A, NP_001272656.1:p.Gly496Asp by direct Sanger sequencing of PCR amplicons as described (Becker *et al.* 2015).

Results

To elucidate the molecular basis and origin of rare coat color patterns in Valais goats, we analyzed whole genome sequence data of representative female goats with the four different coat colors (Figure 2). The black or brown base color phenotype was consistent with the genotypes at the previously reported *TYRP1*:p.Gly496Asp variant (Becker et al. 2015; Dietrich et al. 2015). The Blackneck and the Copperneck goat both carried the wildtpye *ASIP* allele (A^{bz}) in a homozygous state.

Previously described mutant *ASIP* alleles were present in the white Capara Sempione and the greyish Grüenochte. The white Capra Sempione carried one copy of the A^{Wt} allele, which is commonly found in Saanen and Appenzell goats. A^{Wt} involves a triplication of ~150 kp comprising the coding sequence of the *ASIP*, *AHCY* and *ITCH* genes (Fontanesi *et al.* 2009; Henkel *et al.* 2019). Closer inspection of the short-read alignments revealed that the sequenced Capra Sempione had acquired additional smaller structural alterations compared to the original A^{Wt} allele, which comprised a duplication of 1'053 bp inserted into a 10,194 bp deletion of the original A^{Wt} CNV allele (Supplementary Figure S1).

The greyish Grüenochte goat carried one copy of the A^{Pc} allele, which causes the characteristic black and white pattern of Swiss Peacock goats (Henkel *et al.* 2019). The A^{Pc} allele consists of a central quadruplication of ~45 kb with triplicated adjacent flanking regions comprising ~28 kb and ~42 kb respectively.

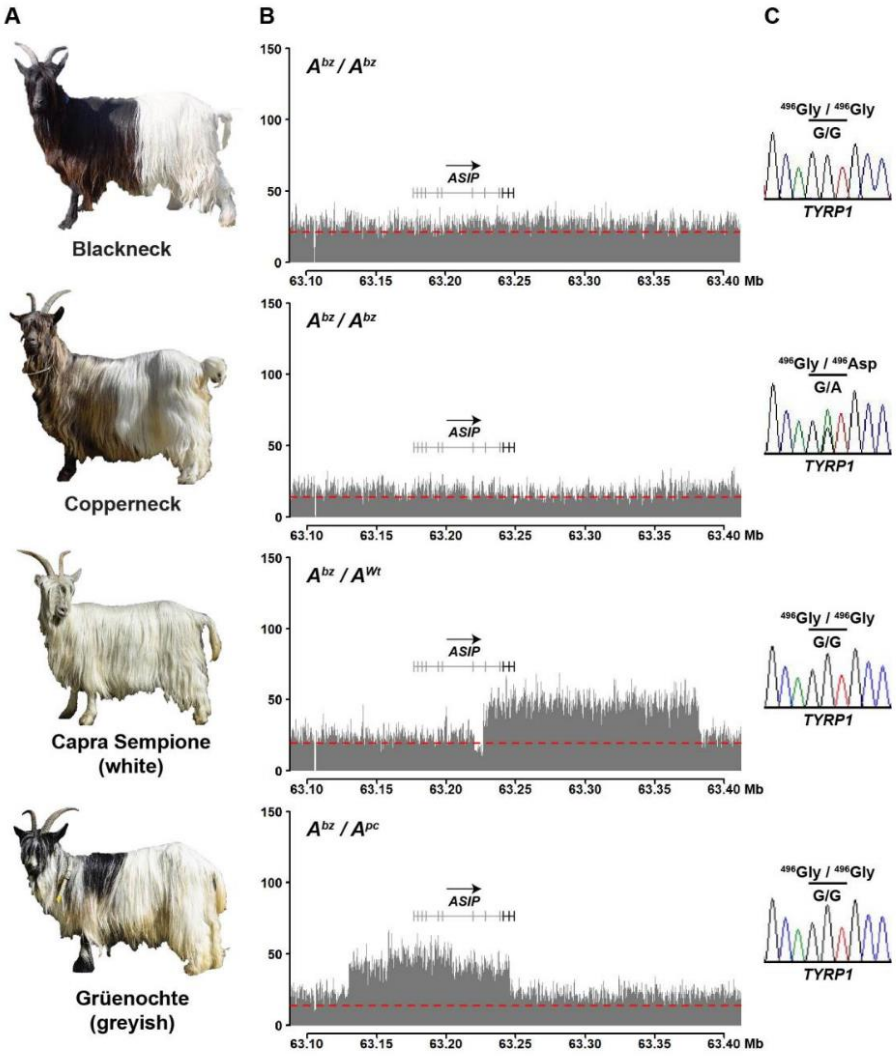

Figure 2. *ASIP* and *TYRP1* variants in Valais goats. **A** Coat color phenotypes of the four varieties of Valais goat. **B** Coverage plots of the *ASIP* locus in Valais goats revealed two distinct CNVs. The Blackneck and Copperneck coverage plot show uniform coverage characteristic for the wild type allele (A^{bz}). The coverage plot of the Sempione and Grüenochte goats revealed CNVs associated with mutant *ASIP* alleles (A^{Wt} and A^{pc}). The dashed red lines indicate the average coverage. **C** Sanger sequencing electropherograms illustrating the variation at *TYRP1*:c.1487G>A (p.Gly496Asp), which causes dominant brown coat color.

To confirm the suspected causal roles of the A^{Wt} and A^{pc} alleles for the white and greyish coat color in Capara Sempione and Grüenochte, resepctively, we designed two diagnostic PCR genotyping assays and genotyped a total of 393 Valais goats for the presence of these two alleles (Table 1; Supplementary Table S1). This experiment revealed a perfect association between the presence of A^{Wt} and the white coat color in Capra Sempione as well as A^{pc} and the greyish coat color in Grüenochte. A^{Wt} and A^{pc} are dominant over the wildtype allele A^{bz}. Breeding records and pedigrees are in agreement with a dominant inheritance of the white and greyish coat color phenotypes over the Blackneck pattern. Our genotyping assay could not distinguish between heterozygous and homozygous mutant animals at the *ASIP* locus. However, the breeding records suggest that a substantial fraction of Capra Sempione and Grüenochte goats have one Blackneck parent and are therefore heterozygous. We did not observe any goat with a heterozygous A^{Wt}/A^{pc} genotype. Based on the known dominance hierarchy of caprine *ASIP* alleles, such a goat would be expected to display the white Capra Sempione phenotype.

Table 1. *ASIP* and *TYRP1* genotype association with specific coat color varieties in Valais goats.

Phenotype	**Blackneck** (n = 266)	**Copperneck** (n = 98)	**Capra Sempione (white)** (n = 13)	**Grüenochte (greyish)** (n = 16)
A^{bz} / A^{bz}	266	98	–	–
A^{Wt} / –	–	–	13	–
A^{pc} / –	–	–	–	16
$TYRP1^{Gly496/Gly496}$	266	–	10	16
$TYRP1^{Gly496/Asp496}$	–	87	3	–
$TYRP1^{Asp496/Asp496}$	–	11	–	–

Discussion

The present study clarified the molecular basis for the four different coat color varieties in Valais goats. Previous research had established that the difference in black or brown base coat color is due to the *TYRP1*:p.Gly496Asp variant. The dominant *TYRP1*^{Asp496} allele resulting in the synthesis of brown eumelanin was most likely introgressed from Toggenburger goats (Becker *et al.* 2015, Dietrich *et al.* 2015). The additional variation resulting in the white or greyish coat color is due to altered pigment type switching caused by mutant *ASIP* alleles. While Blackneck and Copperneck goats exclusively express eumelanin in their pigmented cranial body half, the white Capra Sempione may be assumed to snythesize a very light phaeomelanin in the cranial half of the body, similar to what is seen in white Saanen or Appenzell goats. The "greyish" coat color of the Grüenochte closely resembles the pattern of Peacock goats in the cranial body half, while the caudal body half is devoid of melanocytes as in the other varieties of Valais goats. The similiarity of the pattern in Grüenochte and Peacock goats had been difficult to recognize due to the extremely long hair of Valais goats and the fact that the pigmentation of Valais goats show a highly variable degree of bleaching with age and/or sunlight exposure.

The molecular cause for the depigmentation of the caudal half of all Valais goats is still unknown. Further research is required to elucidate the molecular genetics of this striking pattern.

Our study identified introgression events of two distinct *ASIP* alleles into Valais goats. We observed a variant A^{Wt} allele in the sequenced Capra Sempione. We speculate that this variant A^{Wt} allele is carried by all Capra Sempione but it might also represent a private variation of the sequenced animal. The observed genotype distribution and available pedigree data indicated an autosomal dominant mode of inheritence. These findings support what is already known about the copy number variants at the agouti locus of goat and sheep (Adalsteinsson *et al.* 1994; Norris *et al.* 2008; Fontanesi *et al.* 2009; Henkel *et al.* 2019).

The origin of the A^{Wt} allele is most likely in the Swiss Appenzell or Saanen goats, as both are nearly fixed for this allele, which is responsible for their characteristic white coat color (Fontanesi *et al.* 2009; Henkel *et al.* 2019). The A^{Pc} allele was presumably introduced from Peacock goats as it gives

them their characteristic black and white pattern (Henkel *et al.* 2019). The Swiss goat herding practices represent a potential mediator for the observed introgression. During summer, goats from different farms and breeds are jointly kept on pastures in the mountains, the so-called Alps. This offers a contact zone for the normally closed Swiss goat breeds.

Introgression events of domestic alleles into wild animals are known in various mamalian species and were for example observed in *Capra ibex* (Grossen *et al.* 2014), mouflons (Feulner *et al.* 2013) or in North American gray wolves (Anderson *et al.* 2009). The introgression of the two mutant *ASIP* alleles into Valais goats probably happened before the revision of the breed standard 1938, as anecdotal reports of the Valais goat population suggested a phenotypically quite diverse population during the first half of the 20th century. Offspring of such hybridizations between Appenzell/Saanen and Blackneck or Peacock and Blackneck may have been of particular interest to their owners and remained in the Valais goat population until the revision of the breed standard in 1938, when the breed standard was refined to exclusively allow animals with the Blackneck phenotype.

The presented results show a previously unsuspected introgression of different *ASIP* alleles into Valais goats and provide a molecular characterization of historic breed developments in Switzerland.

Funding

This work was supported by grant 31003A_172964 from the Swiss National Science Foundation.

Acknowledgments

The authors are grateful to all goat owners and breeding organizations who donated samples and shared pedigree data and phenotype information of their animals. We thank Nathalie Besuchet Schmutz and Carmen Rodriguez for expert technical assistance, the Next Generation Sequencing Platform of the University of Bern for performing the high-throughput sequencing experiments, and the Interfaculty Bioinformatics Unit of the University of Bern for providing high performance

computing infrastructure. Furthermore, we thank Christian Gazzarin for the photos of the different Valais goats.

Data Availability

We have deposited the primary data underlying these analyses as follows:

NCBI SRA: SRX1560776, SRX5250095, SRX7715598 and SRX9449177 (see Supplementary Table S2 for more details)

References

Adalsteinsson S, Sponenberg DP, Alexieva S, Russel AJ. 1994. Inheritance of goat coat colors. J Hered.85: 267-272.

Anderson TM, vonHoldt BM, Candille SI, Musiani M, Greco C, Stahler DR et al. 2009. Molecular and evolutionary history of melanism in North American gray wolves. Science.323:1339– 1343.

Andersson L. 2013. Molecular consequences of animal breeding. Curr Opin Genet Dev.23:295-301.

Andersson L. 2016. Domestic animals as models for biomedical research. Ups J Med Sci.121:1-11.

Becker D, Otto M, Ammann P, Keller I, Drögemüller C, Leeb T. 2015. The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the *TYRP1* locus on chromosome 8. Anim Genet.46:50-54.

Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. Nature Genet.49: 643-650.

Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, et al. 2016. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. Anim Genet.47: 727-739.

Dietrich J, Menzi F, Ammann P, Drögemüller C, Leeb T. 2015. A breeding experiment confirms the dominant mode of inheritance of the brown coat colour associated with the ^{496}Asp *TYRP1* allele in goats. Anim Genet.46: 587-588.

FAO. The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture, edited by B.D. Scherf & D. Pilling. 2015. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome. (available at http://www.fao.org/3/a-i4787e/index.html)

Feulner PGD, Gratten J, Kijas JW, Visscher PM, Pemberton JM, Slate J. 2013. Introgression and the fate of domesticated genes in a wild mammal population. Mol Ecol.22:4210–4221.

Fontanesi L, Beretti F, Riggio V, Gomez Gonzalez E, Dall'Olio S, Davoli R, et al. 2009. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. Cytogenet Genome Res.126:333-347.

Glowatzki-Mullis M.L., Muntwyler J., Bäumle E., Gaillard C. 2008. Genetic diversity measures of Swiss goat breeds as decision-making support for conservation policy. Small Ruminant Research 74, 202–11.

Grossen C, Keller L, Biebach I, The International Goat Genome Consortium, Croll D. 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. PLoS Genet.10: e1004438.

Henkel J., Saif R., Jagannathan V., Schmocker C., Zeindler F., Bangerter E. et al. 2019. Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. PLoS Genet.15:e1008536

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics.25:2078-2079.

Menzi F, Keller I, Reber I, Beck J, Brenig B, Schutz E, et al. 2016. Genomic amplification of the caprine *EDNRA* locus might lead to a dose dependent loss of pigmentation. Sci Rep.6: 28438.

Norris BJ, Whan VA. 2008. A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep. Genome research.18(8):1282-93.

R Core Team. R. 2018. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinf.14: 178-192.

Supplementary Material

Supplementary Figure 1. Details of the A^{Wt} allele in the sequenced Capra Sempione.

Supplementary Table 1. Individual goat samples and their genotypes.

Supplementary Table 2. Sequenced goat genomes.

Supplementary Table 3. Primer sequences.

4.3 A deletion spanning the promoter and first exon of the hair cycle-specific *ASIP* transcript isoform in black and tan rabbits

Journal: Animal Genetics
Manuscript status: published
Contributions: participation in genetic experiments, review and editing of manuscript

Displayed version: Accepted
DOI: 10.1111/age.12881

Short communication

A deletion spanning the promoter and first exon of the hair cycle-specific *ASIP* transcript isoform in black and tan rabbits

Anna Letko[1], Beatrice Ammann[1], Vidhya Jagannathan[1,2], Jan Henkel[1,2], Fabienne Leuthard[1,2], Claude Schelling[3], Miguel Carneiro[4,5], Cord Drögemüller[1,2], Tosso Leeb[1,2]

[1] Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland

[2] Dermfocus, University of Bern, 3001 Bern, Switzerland

[3] Clinic for Reproductive Medicine and Center of Clinical Studies, Vetsuisse Faculty, University of Zurich, 8315 Lindau, Switzerland

[4] CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão, Portugal

[5] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

Running title: Black and tan rabbits

Address for correspondence

Tosso Leeb
Institute of Genetics
Vetsuisse Faculty
University of Bern
Bremgartenstrasse 109a
3001 Bern
Switzerland

Phone: +41-31-6312326
E-mail: tosso.leeb@vetsuisse.unibe.ch

Summary

Black and tan animals have tan-coloured ventral body surfaces separated by sharp boundaries from black-coloured dorsal body surfaces. In the a^t mouse mutant, a retroviral 6 kb insertion located in the hair cycle-specific promoter of the murine *Asip* gene encoding agouti signalling protein, causes the black and tan phenotype. In rabbits, three *ASIP* alleles are thought to exist including an a^t allele causing a black and tan coat colour that closely resembles the mouse black and tan phenotype. The goal of our study was to identify the functional genetic variant causing the rabbit a^t allele. We performed a whole genome sequence-based comparative analysis of the *ASIP* gene in one black and tan and three wildtype agouti-coloured rabbits. The analysis identified 75 a^t-associated variants including an 11 kb deletion. The deletion is located in the region of the hair cycle-specific *ASIP* promoter and thus in a region homologous to the site of the retroviral insertion causing the a^t allele in mice. We observed perfect association of the genotypes at this deletion with the coat colour phenotype in 49 rabbits. The comparative analysis and the previous knowledge about the regulation of *ASIP* expression suggest that the 11 kb deletion is the most likely causative variant for the black and tan phenotype in rabbits.

Keywords *Oryctolagus cuniculus*, pigmentation, coat colour, non-coding, whole-genome sequence, promoter, structural variant

Variation in coat colour is governed by numerous genetic loci that influence the production and/or distribution of pigments (Kaelin and Barsh, 2013). Pigment in mammals is exclusively produced by melanocytes that synthesize two types of melanin. These are yellow to red pheomelanin and dark brown to black eumelanin. The basic coat colour in mammals is determined by the ratio of these two pigment types and regulated by the so-called pigment type switching, an intensively studied signalling process (Barsh, 1996; Cieslak *et al.* 2011).

The melanocortin 1 receptor (MC1R) is expressed in the plasma membrane of melanocytes and can bind two ligands: **α-melanocyte stimulating hormone (**α-MSH) and agouti signalling protein (ASIP). α-MSH binding leads to activation of MC1R and promotes eumelanin synthesis via the upregulation of cAMP signaling. ASIP is a competitive inhibitor of α-MSH and its binding to MC1R leads to the production of pheomelanin (Barsh *et al.* 2000). ASIP, therefore, plays an important role in the spatial and temporal control of eumelanin and pheomelanin synthesis (Cieslak *et al.* 2011; Kaelin and Barsh, 2013). The use of different alternative promoters of the *ASIP* gene governs the pigment type-switching patterns, as described in mice (Vrieling *et al.* 1994), pigs (Drögemüller *et al.* 2006) and rabbits (Fontanesi *et al.* 2010).

The murine *Asip* gene contains four alternatively used, untranslated exons **at its 5**'-end, which are under the control of the so-called ventral-specific promoter (exons **1A and 1A')** or the hair cycle-specific promoter, which is located further downstream (exons 1B and 1C). Regulated *Asip* expression synchronized with different stages of the hair cycle is responsible for the banded pigmentation in the hairs of *agouti* mice (Vrieling *et al.* 1994). The black and tan phenotype in the a^t mouse mutant is caused by a ~6 kb retroviral-like insertion in the region of the hair cycle-specific promoter (Bultman *et al.* 1994). In black and tan a^t mice, hairs are no longer banded and show a uniformly yellow or uniformly black pigmentation.

In rabbits, three functional *ASIP* alleles with a dominance hierarchy of $A > a^t > a$ have been described (Fontanesi *et al.* 2010). The dominant *A* allele leads to the wildtype or agouti phenotype with three colour zones visible in the individually banded hairs (Fig. 1). The recessive non-agouti *a* allele caused by a frameshift insertion, NM_001122939.1:c.5_6insA,

results in a black fur (Fontanesi *et al.* 2010). The aim of the present study was to identify the causative variant for the a^t allele in black and tan rabbits that lack banding of the hairs, similar to a^t mice (Fig. 1).

We performed whole-genome sequencing of one black and tan and three wildtype agouti-coloured rabbits (ENA accessions are given in Table S1). The resulting fastq-files were mapped to the rabbit reference genome assembly OryCun2.0 (Carneiro *et al.* 2014) and single nucleotide and small indel variants were called. NCBI annotation release 102 was used to predict their functional effects as described previously (Jagannathan *et al.* 2019). The IGV software (Thorvaldsdottir *et al.* 2013) was used for visual inspection of the regions of interest and identification of structural variants.

The black and tan rabbit was homozygous for the alternative allele at 551 variants in the genomic segment containing the *ASIP* gene (NC_013672.1:g.5,423,362-5,631,747). These comprised 395 SNVs, 155 small indels and one structural variant (Table S2). Filtering of those variants, for which at least one of the control genomes was also homozygous for the alternative allele, reduced the list to 49 associated SNVs, 25 small indels, and the structural variant (Table S2). The structural variant represented an ~11 kb deletion (NC_013672.1:g.5,455,408_5,466,123del; Fig. 2). In rabbit, there are currently two *ASIP* transcript isoforms annotated (NCBI annotation release 102). The deletion removes the entire **first 5'**-untranslated exon of one of these transcripts (NM_001122939.1). Extrapolating from the organization of the murine *Asip* gene, this corresponds to the first exon of the hair cycle-specific transcript.

We confirmed the deletion by PCR followed by Sanger sequencing and genotyped 49 rabbits with different coat colours (Table S1). Out of 19 black and tan rabbits analysed, 17 carried the deletion in a homozygous state. The remaining two black and tan rabbits were heterozygous for the deletion but also heterozygous for the single base insertion causing recessive black and thus presumably compound heterozygous a^t/a. As the *a* allele is recessive to a^t, rabbits with an a^t/a genotype are phenotypically black and tan. None of the 30 rabbits

that were not black and tan carried the deletion in a homozygous state (Table S1). These genotyping results are therefore compatible with a causative role for this deletion.

We note, however, that the rabbit genome assembly still contains some small gaps in the region of the *ASIP* gene. If the causative variant for black and tan is located in a gap of the genome reference assembly or if it is located outside of the genomic interval defined by the flanking *EIF2S2* and *AHCY* genes, we would have missed it in our analysis.

In conclusion, based on comprehensive whole-genome sequencing data we report 75 variants whose genotype distribution was compatible with a causative role in the black and tan phenotype. None of these variants affected the coding sequence of the *ASIP* gene. It is very intriguing that the only associated structural variant in rabbits, an ~11 kb deletion, was in the same region of the *ASIP* gene as the retroviral insertion in the murine a^t allele (Fig. 2). The deletion removed the transcription start site and the first untranslated exon of the presumable hair cycle-specific transcript isoform suggesting that this is the most likely causative variant for the black and tan phenotype. Definitive proof of the causality will require further experiments, e.g. by introducing the rabbit deletion into an agouti (wildtype) genetic background via CRISPR/Cas9-mediated genome editing or by performing comprehensive transcript analyses in black and tan rabbits.

Acknowledgments

The Next Generation Sequencing Platform of the University of Bern is acknowledged for performing the whole genome re-sequencing experiment and the Interfaculty Bioinformatics Unit of the University of Bern for providing computational infrastructure. Regina Regitz is acknowledged for valuable discussions and donating rabbit samples and photos. We thank Sonja Hofstetter for providing the illustrations of a black and tan rabbit, Simon König for the photographs, and all owners and breeders for providing samples, photos and additional data of their animals. MC was supported by the Fundação para a Ciência e Tecnologia (FCT) through POPH-QREN funds from the European Social Fund and Portuguese MCTES (FCT Investigator grant to MC [IF/00283/2014/CP1256/CT0012]).

References

Barsh G.S. (1996) The genetics of pigmentation: from fancy genes to complex traits. *Trends in Genetics* **12**, 299-305.

Barsh G., Gunn T., He L., Schlossman S. & Duke-Cohan J. (2000) Biochemical and genetic studies of pigment-type switching. *Pigment Cell Research* **13 Suppl 8**, 48-53.

Bultman S.J., Klebig M.L., Michaud E.J., Sweet H.O., Davisson M.T. & Woychik R.P. (1994) Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (at) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. *Genes & Development* **8**, 481-90.

Carneiro M., Rubin C.J., Di Palma F. *et al.* (2014) Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074-9.

Cieslak M., Reissmann M., Hofreiter M. & Ludwig, A. (2011). Colours of domestication. *Biological Reviews* **86**, 885-99.

Drögemüller C., Giese A., Martins-Wess F., Wiedemann S., Andersson L., Brenig B., Fries R. & Leeb T. (2006). The mutation causing the black-and-tan pigmentation phenotype of Mangalitza pigs maps to the porcine ASIP locus but does not affect its coding sequence. *Mammalian Genome* **17**, 58-66.

Fontanesi L., Forestier L., Allain D., Scotti E., Beretti F., Deretz-Picoulet S., Pecchioli E., Vernesi C., Robinson T.J., Malaney J.L., Russo V. & Oulmouden A. (2010). Characterization of the rabbit agouti signaling protein (*ASIP*) gene: Transcripts and phylogenetic analyses and identification of the causative mutation of the nonagouti black coat colour. *Genomics* **95**, 166–75.

Jagannathan V., Drögemüller C., Leeb T. & Dog Biomedical Variant Database Consortium (DBVDC) (2019) A comprehensive biomedical variant catalogue based on whole genome sequences of 584 dogs and 6 wolves. *Animal Genetics* doi:10.1111/age.12834.

Kaelin C. & Barsh G. (2013). Genetics of Pigmentation in Dogs and Cats. *Annual Review of Animal Biosciences* **1**, 125-56.

Thorvaldsdóttir H., Robinson J. & Mesirov J. (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-92.

Vrieling H., Duhl D., Millar S., Miller K. & Barsh G. (1994). Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proceedings of the National Academy of Sciences* **91**, 5667-71.

Figure legends

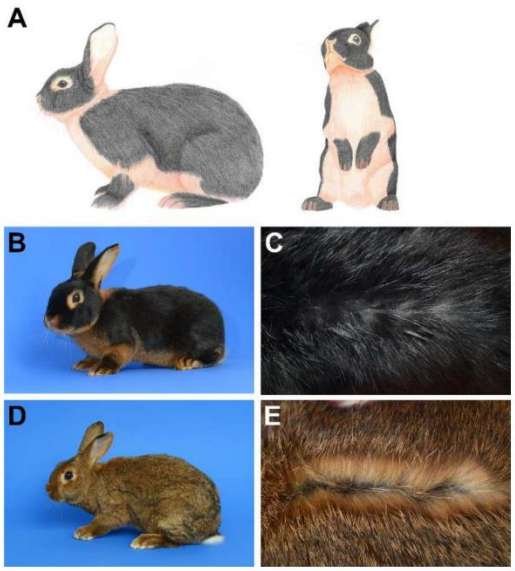

Figure 1 Black and tan coat colour phenotype. (A) Schematic representation of a black and tan rabbit. (B) Photo of a black and tan rabbit. (C) Close-up view of the coat on the back of the same black and tan rabbit. The hair on this rabbit does not show any banding. It is uniformly black on the back and uniformly yellow in the tan areas. (D) Photo of a rabbit with the agouti (wildtype) coat colour. (E) Close-up view of the coat on the back of the same agouti rabbit. The hairs were parted to demonstrate the banding of the individual hairs.

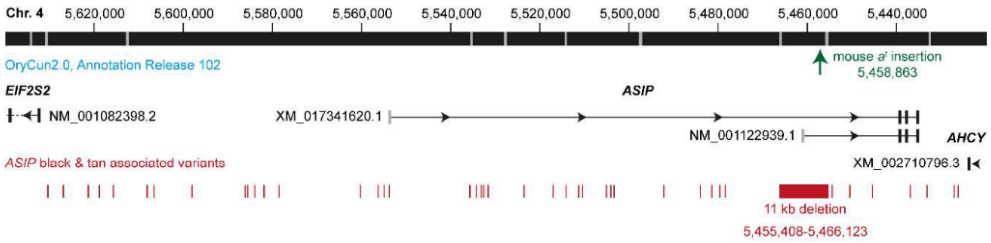

Figure 2 Genomic organization of the rabbit *ASIP* locus and location of variants associated with the black and tan phenotype. On the top, chromosome 4 is represented by a solid black bar. Grey areas represent gaps in the genome reference assembly. The position corresponding to the insertion site of the 6 kb retroviral-like sequence in the murine $a^{t\text{-}2Gso}$ allele is indicated with a dark green arrow (Bultman et al. 1994). The NCBI annotation of the region is indicated. The two annotated *ASIP* transcripts correspond to transcripts 1A and 1C reported previously (Fontanesi *et al.* 2010). We considered variants in the interval between *EIF2S2* and *AHCY* as candidate variants for the black and tan phenotype. The positions of 75 variants that were private to the black and tan rabbit and absent from four agouti-coloured rabbits are indicated in red.

Supporting Information

Table S1 Detailed information about studied rabbits.

Table S2 List of variants of a sequenced black and tan rabbit in the *ASIP* region.

4.4 A novel *KIT* deletion variant in a German Riding Pony with white-spotting coat colour phenotype

Journal:	Animal Genetics
Manuscript status:	published
Contributions:	participation in genetic analyses, review and editing of manuscript
Displayed version:	Accepted
DOI:	10.1111/age.12840

Short Communication

A novel *KIT* deletion variant in a German Riding Pony with white spotting coat colour phenotype

Petra Hug[1,2], Rony Jude[3], Jan Henkel[1,2], Vidhya Jagannathan[1,2], Tosso Leeb[1,2]

[1] Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland

[2] DermFocus, University of Bern, 3001 Bern, Switzerland

[3] RJC, 53919 Weilerswist, Germany

Running title: *KIT* deletion in a German Riding Pony

Address for correspondence

Tosso Leeb
Institute of Genetics
Vetsuisse Faculty
University of Bern
Bremgartenstrasse 109a
3001 Bern
Switzerland

Phone: +41-31-6312326
Fax: +41-31-6312640
E-mail: Tosso.Leeb@vetsuisse.unibe.ch

Summary

White spotting phenotypes in horses may be caused by developmental alterations impairing melanoblast differentiation, survival, migration and/or proliferation. Candidate genes for white spotting phenotypes in horses include *EDNRB*, *KIT*, *MITF*, *PAX3,* and *TRPM1*. We investigated a German Riding Pony with a sabino-like phenotype involving extensive white spots on the body together with large white markings on the head and almost completely white legs. We obtained whole genome sequence data from this horse. The analysis revealed a heterozygous 1273 bp deletion spanning parts of intron 2 and exon 3 of the equine *KIT* gene (Chr3:79,579,925_79,581,197). We confirmed the breakpoints of the deletion by PCR and Sanger sequencing. The knowledge on the functional impact of similar *KIT* variants in horses and other species suggests that this deletion represents a plausible candidate causative variant for the white spotting phenotype. We propose the designation *W28* for the mutant allele.

Keywords: *Equus caballus*; melanocyte; skin; pigmentation; coat colour; structural variant; whole genome sequencing

Melanoblast development, migration, survival and proliferation as well as the full differentiation into mature melanocytes are intricately regulated processes. White spotting phenotypes are the result of an altered embryonic development of the neural crest-derived melanocyte lineage (Thomas & Erickson, 2008). Candidate genes for such phenotypes in horses include *EDNRB*, *KIT*, *MITF*, *PAX3,* and *TRPM1* (OMIA 000629, 000209, 001688, 000214, 001341).

We investigated a one-year-old German Riding Pony with a sabino-like white spotting phenotype. The pony had extensive white body spots, and an almost completely white head and legs (Figure 1). Neither the father nor the mother showed a similar phenotype suggesting a potential *de novo* mutation event, which had given rise to a dominant white spotting allele. The pony was tested negative for 28 known white-spotting alleles (*SB-1*, *W1-W27*) in the *KIT* gene (Brooks & Bailey, 2005; Haase et al. 2007; Haase et al. 2009; Haase et al. 2010; Holl et al. 2010; Hauswirth et al. 2013; Haase et al. 2015; Dürig et al. 2017; Holl et al. 2017; Capomaccio et al. 2017; Hoban et al. 2018) as well as for the splashed white alleles (SW1-3) in the *MITF* and *PAX3* genes (Hauswirth et.al. 2012; Table S1).

We collected an EDTA blood sample from the pony and extracted genomic DNA using the Maxwell RSC Blood DNA Kit and a Maxwell RSC instrument. An Illumina TruSeq PCR-free DNA library with 350 bp insert size was prepared. We collected 240,866,812 million 2 x 150 bp read pairs on a NovaSeq 6000 instrument (29x coverage). Sequencing and read mapping to the EquCab 3 reference assembly was performed as previously described (Jagannathan et al. 2019). The sequence data were deposited under study accession PRJEB14779 and sample accession SAMEA5600769 at the European Nucleotide Archive. This analysis failed to reveal any single nucleotide or small indel variants in the functional candidate genes.

Using the Integrative Genomics Viewer (Robinson et al. 2011), we visually inspected the short read alignments for structural variants in the regions of functional candidate genes *EDNRB*, *KIT*, *MITF*, *PAX3,* and *TRPM1*. A 1273 bp deletion spanning parts of intron 2 and exon 3 of the *KIT* gene was detected, Chr3:79,579,925_79,581,197del (Figure 2).

A primer pair for the amplification of the mutant allele with the deletion was designed. Genomic DNA from the pony was amplified using forward primer ATCAGCGACGAACGTCAAGT, reverse primer GTGTCCTTTCCTTGGTGGGT and AmpliTaq Gold 360 Master Mix. The amplicon was treated with shrimp alkaline phosphatase and exonuclease I and sequenced on an ABI 3730 capillary sequencer. The Sanger sequencing data confirmed the breakpoints of the deletion (Figure 2). The deletion allele was not found in 88 genome sequences from genetically diverse horses (Table S2, Jagannathan et al. 2019).

Exon 3 of the equine *KIT* gene spans 282 bp, of which 75 bp are deleted in the mutant allele. **As the deletion also removes the 3'**-splice site of intron 2, splicing of the *KIT* mRNA most

likely is severely altered. Unfortunately, we had no suitable samples for RNA isolation, which would be required to investigate the functional consequences of the genomic deletion on the transcript level. A comparable deletion in the equine *KIT* gene spanning exons 10-13 gives rise to the *W22* allele that was identified in Thoroughbred Horses (Dürig et al. 2017). Heterozygous $KIT^{W22/+}$ horses have a similar white-spotting phenotype as the German Riding Pony described in this report.

In conclusion, based on the extensive knowledge on the functional effect of *KIT* variants, it is highly plausible that the reported deletion spanning parts of intron 2 and exon 3 causes the white spotting phenotype in the investigated German Riding Pony. In line with previously named equine *KIT* alleles, we propose to designate the newly identified allele *W28*.

Acknowledgements

The authors would like to thank the involved horse owner for donating samples and pictures and for sharing pedigree information. The authors also wish to thank Nathalie Besuchet and Sabrina Schenk for expert technical assistance. The Next Generation Sequencing Platform of the University of Bern is acknowledged for performing the whole genome re-sequencing experiments and the Interfaculty Bioinformatics Unit of the University of Bern for providing high performance computing infrastructure. This study was funded by the Swiss National Science Foundation (31003A_172964). Conflicts of interest: Rony Jude is an equine consultant.

References

Brooks S. A. & Bailey E. (2005) Exon skipping in the *KIT* gene causes a Sabino spotting pattern in horses. *Mammalian Genome* **16**, 893-902.

Capomaccio, S., Milanesi, M., Nocelli, C., Giontella, A., Verini-Supplizi, A., Branca, M., Silvestrelli, M. & Cappelli, K. (2017) Splicing site disruption in the *KIT* gene as strong candidate for white dominant phenotype in an Italian Trotter. *Animal Genetics* **48**, 727-7.

Dürig N., Jude R., Holl H., Brooks S.A., Lafayette C., Jagannathan V., Leeb T. Whole genome sequencing reveals a novel deletion variant in the *KIT* gene in horses with white spotted coat colour phenotypes. *Animal Genetics* **48**, 183-5.

Haase B., Brooks S.A., Schlumbaum A., Azor P.J., Bailey E., Alaeddine F., Mevissen M., Burger D., Poncet P.-A., Rieder S. & Leeb T. (2007) Allelic heterogeneity at the equine *KIT* locus in dominant white (*W*) horses. *PLoS Genetics* **3**, e195.

Haase B., Brooks S.A., Tozaki T., Burger D., Poncet P.-A., Rieder S., Hasegawa T., Penedo C. & Leeb T. (2009) Seven novel *KIT* mutations in horses with white coat colour phenotypes. *Animal Genetics* **40**, 623-9.

Haase B., Obexer-Ruff G., Dolf G., Rieder S., Burger D., Poncet P.A., Gerber V,. Howard J., Leeb T. (2010) Haematological parameters are normal in dominant white Franches-Montagnes horses carrying a KIT mutation. *The Veterinary Journal* **184**, 315-7.

Haase B., Jagannathan V., Rieder S. & Leeb T. (2015) A novel *KIT* variant in an Icelandic horse with white-spotted coat colour. *Animal Genetics*, **46**, 466.

Hauswirth R., Haase B., Blatter M., Brooks S.A., Burger D., Drögemüller C., Gerber V., Henke D., Janda J., Jude R., Magdesian K.G., Matthews J.M., Poncet P.-A., Svansson V., Tozaki T., Wilkinson-White L., Penedo M.C., Rieder S. & Leeb T. (2012) Mutations in *MITF* and *PAX3* cause "splashed white" and other white spotting phenotypes in horses. *PLoS Genetics* **8**, e1002653.

Hauswirth R., Jude R., Haase B., Bellone R.R., Archer S., Holl H., Brooks S.A., Tozaki T., Penedo M.C.T., Rieder S. & Leeb T. (2013) Novel variants in the *KIT* and *PAX3* genes in horses with white-spotted coat colour phenotypes. *Animal Genetics* **44**, 763-5.

Hoban, R., Castle, K., Hamilton, N. & Haase, B. (2018) Novel *KIT* variants for dominant white in the Australian horse population. *Animal Genetics* **49**, 99-100.

Holl H., Brooks S. & Bailey E. (2010) *De novo* mutation of *KIT* discovered as a result of a non-hereditary white coat colour pattern *Animal Genetics* **41 Suppl 2**, 196-8.

Holl H., Brooks S., Carpenter M.L., Bustamante C.D., Lafayette C. (2017) A novel splice mutation within equine KIT and the W15 allele in the homozygous state lead to all white coat color phenotypes. *Animal Genetics,* **48**, 497-8.

Jagannathan V., Gerber V., Rieder S., Tetens J., Thaller G., Drögemüller C. & Leeb T. (2019) Comprehensive characterization of horse genome variation by whole genome sequencing of 88 horses. *Animal Genetics* **50**, 74-7.

Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., Getz G. & Mesirov J.P. (2011) Integrative genomics viewer. *Nature Biotechnology* **29**, 24-6.

Thomas, A.J. & Erickson, C.A. (2008) The making of a melanocyte: the specification of melanoblasts from the neural crest. *Pigment Cell & Melanoma Research* **21**, 598-610.

Figure 1. White spotting phenotype in the German Riding Pony. The distribution of the unpigmented skin and hairs resembles the sabino spotting pattern.

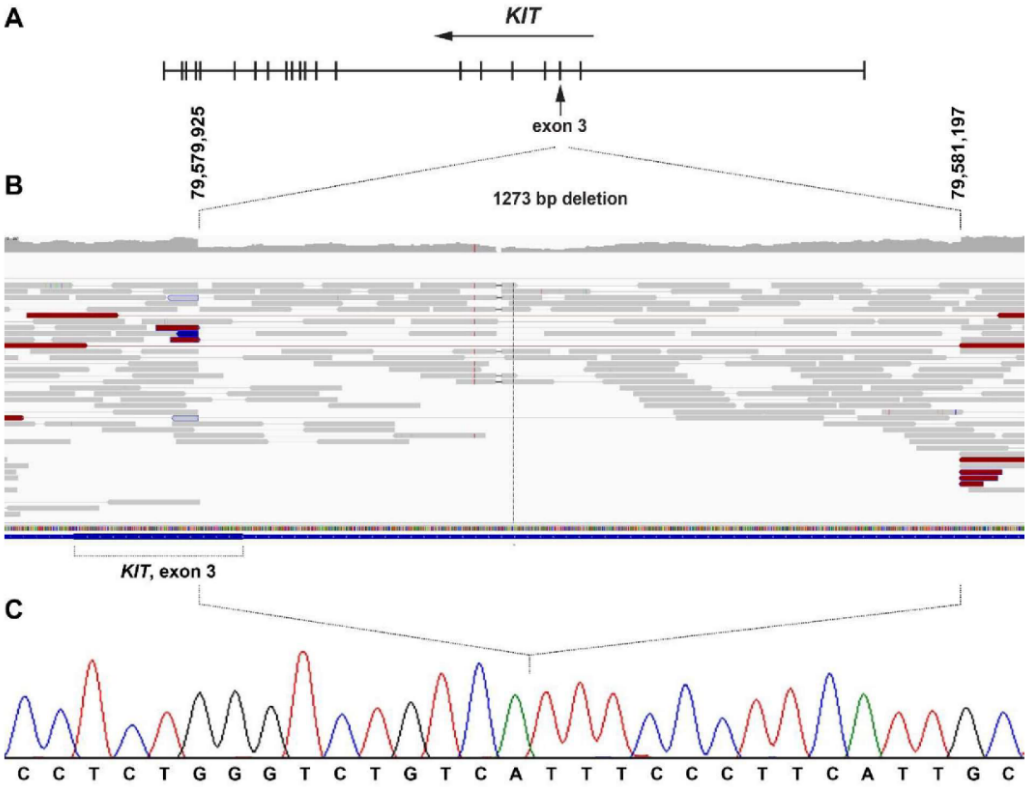

Figure 2. Details of the *KIT* deletion. **(A)** A schematic illustration of the *KIT* gene with its 21 exons. **(B)** IGV screenshot of the illumina short reads in the region of the heterozygous 1273 bp deletion spanning a part of intron 2 and exon 3 of the *KIT* gene. Note the drop in coverage and the truncated read-alignments at the deletion breakpoints. Reads coloured in red indicate read-pairs that map farther apart on the reference genome than the average insert size of the sequencing library. Such read-pairs are indicative for deletions in the sequenced sample. **(C)** Sanger sequence of a PCR amplicon obtained with primers flanking the deletion breakpoints. The deletion can be designated as Chr3:79,579,925_79,581,197del with regards to the EquCab 3 assembly.

Supplementary Material

Table S1. Compilation of variants in candidate genes for white spotting.

Table S2. Accessions of 88 horse genomes and their genotypes at the Chr3:79,579,925_79,581,197del variant.

4.5 Whole-genome sequencing reveals a large deletion in the *MITF* gene in horses with white spotted coat colour and increased risk of deafness

Journal:	Animal Genetics
Manuscript status:	published
Contributions:	genetic analyses, illustrations, original draft, review and editing of manuscript
Displayed version:	Accepted
DOI:	10.1111/age.12762

Short Communication

Whole genome sequencing reveals a large deletion in the *MITF* gene in horses with white spotted coat colour and increased risk of deafness

Jan Henkel[1,2], Christa Lafayette[3], Samantha A. Brooks[4], Katie Martin[3], Laura Patterson-Rosa[4], Deborah Cook[3], Vidhya Jagannathan[1,2], Tosso Leeb[1,2]

[1] Institute of Genetics, Vetsuisse Faculty, University of Bern, 3001 Bern, Switzerland

[2] DermFocus, University of Bern, 3001 Bern, Switzerland

[3] Etalon Inc., Menlo Park, CA 94025, USA

[4] Department of Animal Sciences, University of Florida, Gainesville, FL 32611-0910, USA

Running title: Equine *MITF* deletion

Address for correspondence

Tosso Leeb
Institute of Genetics
Vetsuisse Faculty
University of Bern
Bremgartenstrasse 109a
3001 Bern
Switzerland

Phone: +41-31-6312326
Fax: +41-31-6312640
E-mail: Tosso.Leeb@vetsuisse.unibe.ch

Summary

White spotting phenotypes in horses are highly valued in some breeds. They are quite variable and may range from the common white markings up to completely white horses. *EDNRB*, *KIT*, *MITF*, *PAX3,* and *TRPM1* represent known candidate genes for white spotting phenotypes in horses. For the present study, we investigated an American Paint Horse family segregating a phenotype involving white spotting and blue eyes. Six of eight horses with the white-spotting phenotype were deaf. We obtained whole genome sequence data from an affected horse and specifically searched for structural variants in the known candidate genes. This analysis revealed a heterozygous ~63 kb deletion spanning exons 6-9 of the *MITF* gene (chr16:21,503,211_21,566,617). We confirmed the breakpoints of the deletion by PCR and Sanger sequencing. PCR-based genotyping revealed that all 8 available affected horses from the family carried the deletion. The finding of an *MITF* variant fits well to the syndromic phenotype involving both depigmentation and an increased risk for deafness and corresponds to human Waardenburg syndrome type 2A (WS2A). Our findings will enable more precise genetic testing for depigmentation phenotypes in horses.

Keywords: Equus caballus; melanocyte; pigmentation; coat colour; splashed white; structural variant; heterogeneity

White spotting in horses and other mammals may result from an altered embryonic development of the neural crest melanocyte lineage and a lack of mature melanocytes in the unpigmented **skin areas ("leucism").** Candidate genes for such phenotypes in the horse include *EDNRB*, *KIT*, *MITF, PAX3*, and *TRPM1* (Thomas & Erickson 2008; OMIA 000629, 000209, 001688, 000214, 001341).

The horse currently represents the species with the largest number of molecularly defined white spotting alleles among domesticated animals. A missense variant in the *EDNRB* gene causes the frame overo white spotting pattern or lethal white foal syndrome, if present in heterozygous or homozygous state, respectively (Santschi *et al.* 1998). Variants in *MITF* and *PAX3* cause the so-called splashed white phenotype, which closely resembles human Waardenburg syndrome (Pingault et al. 2010; Hauswirth *et al.* 2012; Hauswirth *et al.* 2013; Dürig et al. 2017). A variant in the equine *TRPM1* gene causes the so-called leopard complex spotting (Bellone *et al.* 2013). Furthermore, according to our knowledge, 29 different functional alleles at the equine *KIT* gene have been described so far. These include the alleles for sabino-1 and tobiano spotting, and an allelic series termed *W1* - *W27* (Brooks & Bailey 2005; Brooks *et al.* 2007; Haase *et al.* 2007; Haase *et al.* 2008; Haase *et al.* 2009; Holl *et al.* 2010; Haase *et al.* 2011; Hauswirth *et al.* 2013; Haase *et al.* 2016; Capomaccio et al. 2017; Holl et al. 2017; Hoban et al. 2018; Table S1).

We studied a family of American Paint horses segregating for a white spotting phenotype resembling the splashed white pattern that could not be explained by any of the previously published white spotting alleles (Figure 1). We had access to samples from one affected stallion and eight of his daughters (Figure S1). Seven of the daughters also had white spotting patterns that could not be explained by their genotypes at the known white spotting variants. One of the daughters had a regular frame overo phenotype, which was inherited from the dam. The coat colour phenotypes in this family were quite variable as all horses carried additional known white-spotting alleles. All affected horses had white faces. The amount of body pigmentation was variable and one of the studied horses was even completely white. At least six of the affected horses had blue eyes. The quality of our photos of the two remaining affected horses did not allow unambiguous determination of their eye colour. Six out of eight horses with unexplained white spotting phenotypes were deaf.

We prepared an Illumina TruSeq PCR free genomic DNA library and re-sequenced the genome of one affected horse at 19x coverage using 2 x 150 bp reads on an Illumina NovaSeq 6000 instrument. Sequencing and read mapping to the EquCab 3 reference assembly was performed as previously described (Jagannathan et al. 2018). Data were deposited at the European Nucleotide Archive (study accession PRJEB14779, sample SAMEA4822838). This analysis failed to reveal any single nucleotide or small indel variants in the candidate genes *EDNRB*, *KIT*, *MITF, PAX3*, and *TRPM1*. In order to search for large structural variants we

visually inspected the read alignments for these genes in the Integrative Genomics Viewer (IGV; Robinson *et al.* 2011).

The analysis revealed a large heterozygous deletion spanning the last four exons of the *MITF* gene corresponding to exons 6 – 9 of the *MITF-M* transcript isoform (Figure 2). Specifically, 63,407 bp were missing (chr16:21,503,211_21,566,617). There is currently no RefSeq entry for the equine melanocyte-specific MITF-M transcript isoform available. We therefore used the accessions JN896378.1 (mRNA) and AFH66983.1 (protein) to analyse the putative effects of the deletion on the transcript and protein. Assuming regular splicing and polyadenylation of the remaining exons, the deletion removes 706 nucleotides (56%) from the open reading frame (JN896378.1:c.555_1260del). This includes the codons required to encode the functionally important bHLH-Zip domain required for DNA binding. We did not find any other obvious structural variants in the *EDNRB*, *KIT, PAX3*, and *TRPM1* genes.

We designed primers flanking the deletion (TTAGCAATAAGCCACTGGTC, TCATTGTGTCCAGGCTGCTG) and confirmed the breakpoints of the deletion by PCR and Sanger sequencing (Figure 2). A PCR with these primers and ATG360 polymerase (ThermoFisher) was used as diagnostic assay to genotype additional horses for the deletion. All eight affected horses from this family, from which genomic DNA was available, carried the deletion. The phenotype resembling the splashed white pattern with an extremely large blaze on the head and frequently associated with blue eyes is similar to the phenotypes of horses with other mutant *MITF* alleles. Most of the horses carrying this deletion were deaf. This confirms earlier observations that horses with a lack of functional *MITF* have an increased risk for deafness (Hauswirth et al. 2012). All deaf horses carried additional white spotting alleles, which may have exacerbated the functional impact of the *MITF* deletion.

In conclusion, our study revealed a large structural variant at the equine *MITF* gene, which most likely causes a splashed white depigmentation phenotype and predisposes to deafness.

Acknowledgements

The authors would like to thank all involved horse owners for donating samples and pictures and for sharing pedigree information of their horses. The authors also wish to thank Nathalie Besuchet, Muriel Fragnière, and Sabrina Schenk for expert technical assistance. The Next Generation Sequencing Platform of the University of Bern is acknowledged for performing the whole genome re-sequencing experiments and the Interfaculty Bioinformatics Unit of the University of Bern for providing high performance computing infrastructure. This study was supported by grant 31003A_172964 from the Swiss National Science Foundation. Conflicts of interest: Christa Lafayette, Katie Martin, and Deborah Cook are affiliated with a genetic testing laboratory offering tests for white spotting in horses.

References

Bellone RR, Holl H, Setaluri V, Devi S, Maddodi N, Archer S, Sandmeyer L, Ludwig A, Foerster D, Pruvost M, Reissmann M, Bortfeldt R, Adelson DL, Lim SL, Nelson J, Haase B, Engensteiner M, Leeb T, Forsyth G, Mienaltowski MJ, Mahadevan P, Hofreiter M, Paijmans JL, Gonzalez-Fortes G, Grahn B, Brooks SA (2013) Evidence for a retroviral insertion in *TRPM1* as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLoS One* **8**, e78280.

Brooks S. A. & Bailey E. (2005) Exon skipping in the *KIT* gene causes a Sabino spotting pattern in horses. *Mammalian Genome* **16**, 893-902.

Brooks S.A., Lear T.L., Adelson D.L. & Bailey E. (2007) A chromosome inversion near the KIT gene and the Tobiano spotting pattern in horses. *Cytogenetic & Genome Research* **119**, 225-30.

Capomaccio S., Milanesi M., Nocelli C., Giontella A., Verini-Supplizi A., Branca M., Silvestrelli M. & Cappelli K. (2017) Splicing site disruption in the *KIT* gene as strong candidate for white dominant phenotype in an Italian Trotter. *Animal Genetics* **48**, 727-8.

Dürig N., Jude R., Jagannathan V. & Leeb T. (2017) A novel *MITF* variant in a white American Standardbred foal. *Animal Genetics* **48**, 123-4.

Haase B., Brooks S.A., Schlumbaum A., Azor P.J., Bailey E., Alaeddine F., Mevissen M., Burger D., Poncet P.-A., Rieder S. & Leeb T. (2007) Allelic heterogeneity at the equine *KIT* locus in dominant white (*W*) horses. *PLoS Genet* **3**, e195.

Haase B., Jude R., Brooks S.A. & Leeb T. (2008) An equine chromosome 3 inversion is associated with the tobiano spotting pattern in German horse breeds. *Animal Genetics* **39**, 306-9.

Haase B., Brooks S.A., Tozaki T., Burger D., Poncet P.-A., Rieder S., Hasegawa T., Penedo C. & Leeb T. (2009) Seven novel *KIT* mutations in horses with white coat colour phenotypes. *Animal Genetics* **40**, 623-9.

Haase B., Rieder S., Tozaki T., Hasegawa T., Penedo M.C., Jude R. & Leeb T. (2011) Five novel *KIT* mutations in horses with white coat colour phenotypes. *Animal Genetics* **42**, 337-9.

Hauswirth R., Haase B., Blatter M., Brooks S.A., Burger D., Drögemüller C., Gerber V., Henke D., Janda J., Jude R., Magdesian K.G., Matthews J.M., Poncet P.-A., Svansson V., Tozaki T., Wilkinson-White L., Penedo M.C., Rieder S. & Leeb T. (2012) Mutations in *MITF* and *PAX3* cause "splashed white" and other white spotting phenotypes in horses. *PLoS Genetics* **8**, e1002653.

Hauswirth R., Jude R., Haase B., Bellone R.R., Archer S., Holl H., Brooks S.A., Tozaki T., Penedo M.C.T., Rieder S. & Leeb T. (2013) Novel variants in the *KIT* and *PAX3* genes in horses with white-spotted coat colour phenotypes. *Animal Genetics* **44**, 763-5.

Hoban R., Castle K., Hamilton N. & Haase B. (2018) Novel *KIT* variants for dominant white in the Australian horse population. *Animal Genetics* **49**, 99-100.

Holl H., Brooks S. & Bailey E. (2010) *De novo* mutation of *KIT* discovered as a result of a non-hereditary white coat colour pattern *Animal Genetics* **41 Suppl 2**, 196-198.

Jagannathan V., Gerber V., Rieder S., Tetens J., Thaller G., Drögemüller C. & Leeb T. (2018) Comprehensive characterization of horse genome variation by whole genome sequencing of 88 horses. *Animal Genetics*, manuscript submitted for publication.

Pingault V., Ente D., Dastot-Le Moal F., Goossens M., Marlin S. & Bondurand N. (2010) Review and update of mutations causing Waardenburg syndrome. *Human Mutation* **31**, 391-406.

Robinson J.T., Thorvaldsdóttir H., Winckler W., Guttman M., Lander E.S., Getz G. & Mesirov J.P. (2011) Integrative genomics viewer. *Nature Biotechnology* **29**, 24-6.

Santschi E.M., Purdy A.K., Valberg S.J., Vrotsos P.D., Kaese H. & Mickelson J.R. (1998) Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mammalian Genome* **9**, 306-9.

Thomas A.J. & Erickson C.A. (2008) The making of a melanocyte: the specification of melanoblasts from the neural crest. *Pigment Cell & Melanoma Research* **21**, 598-610.

Figure 1 White spotting phenotype resembling the splashed white pattern. **(A, B)** This horse had light blue eyes, a white face, white legs and a small white belly spot. The horse was deaf. **(C, D)** This horse had a more pronounced white spotting phenotype compared to its full sibling shown in (A) and (B). It also had light blue eyes and was deaf. Genotypes at important white spotting loci are indicated. *MITF*^{del} designates the new *MITF* allele identified in this study.

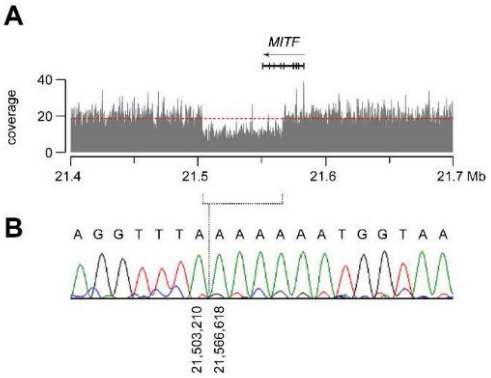

Figure 2 Details of the *MITF* deletion. **(A)** A coverage plot of the whole genome sequence data indicates a heterozygous ~63 kb deletion comprising exons 6-9 of the *MITF* gene. **(B)** Sanger sequencing of a PCR product from the deletion allele precisely defined the breakpoints of the deletion (Chr16:21,503,211_21,566,617del, EquCab 3 assembly).

Supplementary Material:

Figure S1 Pedigree of the studied horses.

Table S1 Compilation of variants in candidate genes for white spotting.

5. Discussion and perspective

During my PhD studies, I contributed to the identification of candidate causative genetic variants of various coat color phenotypes in a range of domestic animals. The investigated phenotypes could be roughly clustered into two groups. In the first group are rare coat color mutants with potential pleiotropic effects. The second group comprises breed specific coat colors with no known pleiotropic effects. The first group involves a white spotting phenotype in an American Paint horse family with increased risk of deafness and a white spotting phenotype in a young German Riding Pony. The second group contains the analysis of breed-defining coat color phenotypes. It includes the calculation of selection signatures in a number of goat breeds, the analysis of different coat color phenotypes segregating in one particular Swiss goat breed and investigations of the black and tan coat color phenotype in rabbits. For the exploration of the above mentioned phenotypes, mainly two different genetic approaches were applied, a candidate gene approach relying on a relatively small sample size and the calculation of selection signatures facilitated by the power of a large comprehensive screen. Interestingly all of the examined phenotypes revealed some sort of structural variation potentially disrupting gene function or altering gene expression.

Of the first group, we received samples from an American Paint horse family displaying variable coat color phenotypes as all of them carried known white spotting alleles (Overview Table A1). Nonetheless, the detected alleles could not explain the observed phenotype completely, nor provided they an explanation for six of the eight horses being deaf. We sequenced one of the horses. After visually going through the list of candidate genes a large heterozygous deletion, spanning four exons of *MITF* was discovered, likely representing a complete loss of function allele. A lack of functional MITF as cause for an increased risk for deafness was already reported in another study in horses [51] and could be confirmed by my study. The additional white spotting alleles might have exacerbated the functional impact of the *MITF* deletion. This study indicates what devastating effects genetic variants of *MITF* can have, highlighting its important function during melanocyte development. The discovered variant is now known as SW5 (splashed white 5) by diagnostic laboratories offering genetic testing. Genetic testing is offered as breeding of horses carrying the variant should be avoided due to animal welfare concerns.

Analysis of a young German Riding Pony with a white spotting phenotype revealed a *de novo* mutation event. Neither father nor mother showed a similar phenotype. Similar to the American Paint horse family, we could not detect any likely causative SNV or small indel and therefore searched for structural variants in the candidate genes. A deletion at *KIT* was identified,

removing the 3' splice site of intron 2, most likely altering the splicing of the *KIT* mRNA resulting in this interesting phenotype. The discovered variant was termed W28 but it is unlikely that this variant will be frequent in horses as it arose due to a *de novo* mutation event and is probably restricted to the analyzed German Riding Pony. This variant together with other dominant white causing variants in horses (Table A1) highlight the role of KIT during melanocyte development as defects influence melanocyte survival [31–33].

This could be further shown by the findings of large complex structural variations downstream of *KIT* potentially influencing its expression in two Pakistani goat breeds. *KIT* requires precise regulation of its temporal and spatial expression and disruptions lead to various coat color phenotypes. The gene is flanked by several hundred kilobases of non-coding genomic DNA on either side harboring important regulatory elements as shown by multiple studies [35,156,172–174]. All found and discussed phenotypes caused by mutant *KIT* in this thesis are characterized by striking alterations in pigmentation. The observed non-coding alteration in goats did not lead to any obvious pleiotropic effects on other *KIT* dependent cell types. The W28 variant of the German Riding Pony on the other side is expected to have serious pleiotropic effects in animals that carry the mutant allele in a homozygous state. The predicted complete lack of functional KIT in homozygous mutant animals might be potentially embryonic lethal. Therefore, mating of two horses carrying heterozygous loss of function variants in *KIT* should be avoided due to animal welfare concerns.

Another set of interesting structural variants found by investigating selection signatures harboring coat color candidate genes affected the caprine *ASIP* locus. In contrast to *KIT*, *ASIP* does not contain a large 3'- flanking region and regulatory elements of the *ASIP* gene are mostly located upstream of the coding exons in a large genomic region harboring several alternative promotors [84]. In goats, I could identify complex structural variations of this region affecting and giving rise to multiple non-coding 5'-exons. This high number of non-coding 5'-exons came by surprise as it was not observed in other mammals so far [82,84]. In another study on rabbits, our group identified a large deletion covering the hair cycle specific *ASIP* promotor and resulting in a black and tan phenotype (Figure 1 E). The observed variants have an effect on the temporal and spatial control of pigment-type switching and help to advance our understanding of the complex regulation of transcription at the *ASIP* locus.

Investigation of coat color phenotypes not only revealed the potential causative variants but also brought light on historic breed developments. By analyzing rare coat color patterns of the Swiss Valais Blackneck goat breed, we observed Introgression events of two distinct *ASIP* alleles. The most likely source of the alleles were Saanen/Appenzell and Peacock goats, as they are fixed for the respective *ASIP* alleles. An important mediator of the observed introgression event was potentially the Swiss goat herding practice, the so-called Alps, which

offer a contact zone for the normally closed Swiss goat breeds. As goats from different farms and breeds are jointly kept on pastures in the Swiss mountains, during summer time. This might have had even more relevance in the past, as the breeds were geographically separated [175]. The introgression events are potentially relatively old (~100 years), as anecdotal breeder reports suggest a phenotypically diverse goat population before the consolidation of the breed standard in 1938.

During recent years the research on selection signatures in domestic animals increased. The usage of comprehensive screens yielded many interesting findings regarding domestication and adaptation to different factors. This could be shown by multiple studies conducted by Leif Andersson's group [35,166–168] and others [176–180]. While I focused on coat color phenotypes during my PhD, the underlying data have the potential to answer many more questions. As an example, examinations of the Pakistani goat breeds included in my search for selection screen revealed an interesting height associated variant in goats of medium to large body size [181]. Similar findings could be made in Swiss goat breeds (ongoing work).
The used dataset of whole genome sequenced Swiss goats was made publicly available and is shared with colleagues for future studies on Swiss goat diversity. As part of the VarGoats consortium (http://www.goatgenome.org/vargoats.html) the produced data will help to build a comprehensive catalog of genetic variation in goat genomes related to domestication and adaptation.

The Valais Blackneck goat carries the most interesting coat color phenotype studied during my PhD, which is characterized by a striking depigmentation of the caudal half of the animal. I was able to identify the potential cause for the rare coat colors segregating in the breed. However, unfortunately, the molecular cause for the depigmentation of the caudal half of all Valais goats is still unknown. Ongoing research on this unique phenotype indicate that it might be more complex, potentially involving structural variations. So far no promising private candidate variant nor region could be identified. Due to its complexity a cross breeding experiment was recently started, expected to provide data in 2023. Additionally, a long read sequencing experiment of one Valais Blackneck goat using the third generation sequencing instrument Sequel II was performed. These long read sequencing data together with short read data and results of the cross breeding should help to identify the causative variant.

The proof for causality is a challenging task in genetic analysis as it is highly dependent on the investigated trait or disorder, the penetrance of the identified variant and the available sample size. Sufficient proof of causality is typically achieved for traits with variants exhibiting high penetrance and exerting large effect sizes through genetic linkage analysis in large families following Mendelian patterns of inheritance [182]. To provide robust evidence of causality for

more complex traits often require more sophisticated functional experiments. Such experiments may involve CRISPR/Cas9 enabled targeted genome editing or functional reporter gene studies to confirm postulated regulatory effects on expression as for example shown in Mishra *et al.* 2017 [183].

In conclusion, this thesis showed the great potential of studying interesting coat color phenotypes in domestic animals using large datasets produced with NGS techniques. The identified variants in horses provide the basis for genetic testing, which enables targeted breeding. In case of the identified deleterious *MITF* allele (SW5), breeding should be completely avoided, since carrier individuals have a higher risk of deafness. The identification of regulatory *ASIP* and *KIT* variants can contribute to a better mechanistic understanding on the spatial and temporal regulation of transcription as well as the complex regulatory networks pigmentation.

6. Acknowledgements

First of all, I would like to thank Prof. Dr. Tosso Leeb for being my supervisor and for giving me the opportunity of this work. You always supported and trusted me. Thank you. You're such a great boss and person.

Thank you my to thesis committee; Prof. Dr. Christine Flury for being my co-supervisor and. Dr. Rémy Bruggmann for being my mentor. I would also like to thank Prof. Dr. Gesine Lühken for agreeing to act as my external co-referee.

As another important person of the Institute of Genetics, I would like to thank Prof. Dr. Cord Drögemüller for his support. You always create a nice atmosphere in the group and help with daily issues. You and Tosso complement each other perfectly. Keep it going.

A big thank you to Dr. Vidhya Jagannathan. She taught me the secrets of bioinformatics and always had an open ear for problems. Without her help, my projects would have been much more difficult. It is now again your office.

As the core of the group since I started, I would like to thank my fellow PhD students Anna Letko and Irene Häfliger. It was always fun working with you and organizing stuff. Thanks to all past and present group members. I met so many during my 4 years that it is impossible to name them all. I will miss you.

I want to thank the lab technicians, past and present ones, but especially Sabrina Schenk, Nathalie Besuchet-Schmutz and Muriel Fragnière. Without you, my thesis would not have been possible.

I would like to thank my family and dear friend Markus for always being there for me. Especially thank you to my mother and father for enabling me to study and supporting me as best as possible. Even as home is far, I tried to visit you as often as possible and I loved it every time.

A special thanks for my love Meike. Thank you for supporting me and always being there with me. I cannot wait till corona is over and we can finally marry.

I would also like to thank the Swiss National Science Foundation (SNSF) for funding this project.

7. Curriculum vitae

Removed due to data privacy reasons

8. List of publications

Published articles

Henkel J, Lafayette C, Brooks SA, Martin K, Patterson-Rosa L, Cook D, Jagannathan V, Leeb T. Whole-genome sequencing reveals a large deletion in the *MITF* gene in horses with white spotted coat colour and increased risk of deafness. *Animal Genetics.* 2019 Apr; 50(2): 172-174. doi: 10.1111/age.12762.

Buck JM, Sherman J, Bártulos CR, Serif M, Halder M, **Henkel J**, Falciatore A, Lavaud J, Gorbunov MY, Kroth PG, Falkowski PG, Lepetit B. Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom *Phaeodactylum tricornutum*. *Nature Communications.* 2019 Sep 13; 10(1): 4167. doi: 10.1038/s41467-019-12043-6.

Hug P, Jude R, **Henkel J**, Jagannathan V, Leeb T. A novel *KIT* deletion variant in a German Riding Pony with white-spotting coat colour phenotype. *Animal Genetics.* 2019 Dec; 50(6): 761-763. doi: 10.1111/age.12840.

Henkel J, Saif R, Jagannathan V, Schmocker C, Zeindler F, Bangerter E, Herren U, Posantzis D, Bulut Z, Ammann P, Drögemüller C, Flury C, Leeb T. Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. *PLoS Genetics.* 2019 Dec 16; 15(12): e1008536. doi: 10.1371/journal.pgen.1008536.

Saif R, **Henkel J**, Jagannathan V, Drögemüller C, Flury C, Leeb T. The *LCORL* locus is under selection in large-sized Pakistani goat breeds. *Genes (Basel).* 2020 Feb 5; 11(2): 168. doi: 10.3390/genes11020168.

Letko A, Ammann B, Jagannathan V, **Henkel J**, Leuthard F, Schelling C, Carneiro M, Drögemüller C, Leeb T. A deletion spanning the promoter and first exon of the hair cycle-specific *ASIP* transcript isoform in black and tan rabbits. *Animal Genetics*. 2020 Feb; 51(1): 137-140. doi: 10.1111/age.12881.

Linek M, Doelle M, Leeb T, Bauer A, Leuthard F, **Henkel J,** Bannasch D, Jagannathan V, Welle MM. *ATP2A2* SINE Insertion in an Irish Terrier with Darier Disease and Associated Infundibular Cyst Formation. *Genes (Basel)*. 2020 Apr 28; 11(5): 481. doi: 10.3390/genes11050481.

Unpublished articles

Henkel J, Dubacher A, Bangerter E, Herren U, Ammann P, Drögemüller C, Flury C, Leeb T. Introgression of *ASIP* and *TYRP1* alleles explains coat color variation in Valais goats. Journal of Heredity. *Accepted*

Bannasch DL, Kaelin CB, Letko A, Loechel R, Hug P, Jagannathan V, **Henkel J**, Roosje P, Hytönen MK, Lohi H, Arumilli M, DoGA consortium, Minor KM, Mickelson JR, Drögemüller C, Barsh GS, Leeb T. Dog color patterns explained by modular promoters of ancient canid origin. *In revision*

Conference abstracts

Henkel J, Lafayette C, Brooks SA, Martin K, Patterson-Rosa L, Cook D, Jagannathan V, Leeb T (2019) Eine grosse Deletion im *MITF* Gen in weiss-gescheckten Pferden mit erhöhtem Taubheitsrisiko. Oral presentation at the Netzwerktagung Pferdeforschung in Avenches, Switzerland, 10.04.2019, Agroscope Science. 2019; 161:243.

Henkel J, Saif R, Jagannathan V, Schmocker C, Zeindler F, Bangerter E, Herren U, Posantzis D, Bulut Z, Ammann P, Drögemüller C, Flury C, Leeb T (2019) Selektionssignaturen und neue Erkenntnisse zur Fellfarbe in Schweizer Ziegenrassen. Poster at the Swiss Association for Animal Sciences in Lindau, Switzerland, 16.04.2019.

Letko A, Loechel R, Minor KM, Mickelson JR, Jagannathan V, **Henkel J**, Ammann B, Carneiro M, Drögemüller C, Leeb T (2019) Structural genetic variants near the hair cycle-specific *ASIP* promoter in black and tan dogs. Poster at the 10. International Conference on Canine and Feline Genetics and Genomics in Bern, Switzerland, 26. – 29.05.2019.

Henkel J, Saif R, Jagannathan V, Drögemüller C, Flury C, Leeb T (2019) Selection signatures in goat breeds reveal the molecular basis for six different coat color phenotypes. Oral presentation and poster at the 37. International Society for Animal Genetics Conference in Lleida, Spain, 7. – 12.07.2019.

Henkel J, Saif R, Jagannathan V, Drögemüller C, Flury C, Leeb T (2020) Identification of selection signatures in 20 diverse goat breeds. Oral presentation and poster at the Plant & Animal Genomes XXVIII Meeting in San Diego, CA, U.S.A., 11. – 15.01.2020.

Henkel J, Jagannathan V, Drögemüller C, Flury C, Leeb T (2021) Investigating the genome of a Valais Blackneck goat for structural variants. Oral presentation at the Virtual PacBio Day, 23.02.2021.

9. Appendix

Table A1. Causal variants for white spotting phenotypes in horses.

Gene	Variant cDNA	Variant Protein/RNA	Phenotype	Horse Breed	Allele
EDNRB	NM_001081837.2:c.353-354TC>AG	NP_001157338.1:p.I118K	Frame Overo white spotting (heterozygous), lethal white foal syndrome (homozygous)	Multiple	LWO [23]
KIT	NM_001163866.2:c.338-1G>C	r.spl?	Dominant white	Thoroughbred	W7 [34]
KIT	NM_001163866.2:c.559_563delTCTGC	NP_001157338.1:p.S187DfsX10	Dominant white	Thoroughbred	W12 [184]
KIT	NM_001163866.2:c.338-1197_414del	NP_001157338.1:p.D113_L138del	Dominant white	German Riding Pony	W28 [7]
KIT	NM_001163866.2:c.668T>C	NP_001157338.1:p.L223P	Dominant white	Thoroughbred	W25 [185]
KIT	NM_001163866.2:c.706A>T	NP_001157338.1:p.K236X	Dominant white	Arabian	W3 [186]
KIT	NM_001163866.2:c.756+1G>C	r.spl?	Dominant white	Arabian	W23 [187]
KIT	NM_001163866.2:c.856G>A	NP_001157338.1:p.G286R	Dominant white	Thoroughbred	W6 [34]
KIT	NM_001163866.2:c.1126_1129delGAAC	NP_001157338.1:p.E376FfsX3	Dominant white	Quarter	W10 [34]
KIT	NM_001163866.2:c.1322A>G	NP_001157338.1:p.Y441C	Dominant white	Arabian	W19 [188]
KIT	NM_001163866.2:c.1346+1G>A	r.spl?	Dominant white	Swiss Warmblood	W18 [188]
KIT	NM_001163866.2:c.1473T>G	NP_001157338.1:p.C491W	Dominant white	Thoroughbred	W27 [185]
KIT	NM_001163866.2:c.1597T>C	NP_001157338.1:p.C533R	Dominant white	Arabian	W15 [189]
KIT	NM_001163866.2:c.1789G>A	NP_001157338.1:p.G597R	Dominant white	Holstein	W9 [34]
KIT	NM_001163866.2:c.1805C>T	NP_001157338.1:p.A602V	Dominant white	Camarillo White	W4 [186]
KIT	NM_001163866.2:c.1960G>A	NP_001157338.1:p.G654R	Dominant white	Thoroughbred	W2 [186]
KIT	NM_001163866.2:c.1529_1978del	NP_001157338.1:p.E510_G659del	Dominant white	Thoroughbred	W22 [190]
KIT	NM_001163866.2:c.2001A>T NM_001163866.2:c.2021T>C	NP_001157338.1:p.E667D NP_001157338.1:p.L674P	Dominant white	Japanese Draft	W17 [189]

Table A1 (continued). Causal variants for white spotting phenotypes in horses.

Gene	Variant cDNA	Variant Protein/RNA	Phenotype	Horse Breed	Allele
KIT	NM_001163866.2:c.2021T>A	NP_001157338.1:p.L674H	Dominant white	Berber	W30 [191]
KIT	NM_001163866.2:c.2045G>A	NP_001157338.1:p.R682H	Dominant white	Multiple	W20 [188]
KIT	NM_001163866.2:c.2151C>G	NP_001157338.1:p.Y717X	Dominant white	Franches-Montagnes	W1 [186]
KIT	NM_001163866.2:c.2193delG	NP_001157338.1:p.T732QfsX9	Dominant white	Thoroughbred	W5 [34]
KIT	NM_001163866.2:c.2222-1G>A	r.spl?	Dominant white	Icelandic	W8 [34]
KIT	NM_001163866.2:c.2349+1G>A	r.spl?	Dominant white	Italian Trotter	W24 [192]
KIT	NM_001163866.2:c2350-13A>T	r.spl?	Sabino white spotting	Multiple	SB1 [193]
KIT	NM_001163866.2:c.2369delC	NP_001157338.1:p.A790Efs*20	Dominant white	Icelandic	W21 [194]
KIT	NM_001163866.2:c.2392_2445del	NP_001157338.1:p.H798_N815del	Dominant white	Thoroughbred	W14 [189]
KIT	NM_001163866.2:c.2472+5G>C	r.spl?	Dominant white	Miniature and Quarter	W13 [189]
KIT	NM_001163866.2:c.2489A>T	NP_001157338.1:p.K830I	Dominant white	Oldenburg	W16 [189]
KIT	NM_001163866.2:c.2536delA	NP_001157338.1:p.S846Vfs*15	Dominant white	Thoroughbred	W26 [185]
KIT	NM_001163866.2:c.2684+1G>A	r.spl?	Dominant white	South German Draft	W11 [34]
KIT	Larger chromosomal inversion		Tobiano white spotting	Multiple	TO [156]
MITF	ENSECAT00000006375.3:c.1-179_1-178insTAGGTTATTAT		Splashed white, Risk for deafness?	Multiple	SW1 [51]
MITF	ENSECAT00000006375.3:c.519_523del GTGTC	ENSECAP00000004506.1:p.C174 Sfs*20	Splashed white, Risk for deafness?	Quarter	SW3 [51]
MITF	ENSECAT00000006375.3:c.560-1569_1260del	ENSECAP00000004506.1:p.A187_C419del	Splashed white, increased risk for deafness	Quarter and Paint	SW5 [6]
MITF	ENSECAT00000006375.3:c.711-3228_1260del	ENSECAP00000004506.1:p.D237_C419del	Splashed white, Risk of deafness?	Quarter and Paint	SW6 [53]
PAX3	XM_001495160.1:c.95C>G	XP_001495210.1:p.P32R	Splashed white	Appaloosa	SW4 [188]
PAX3	XM_001495160.1:c.209G>A	XP_001495210.1:p.C70Y	Splashed white	Multiple	SW2 [51]

10. Figure index

11. References

1. Hoekstra HE. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity (Edinb)*. 2006; 97: 222–234. doi:10.1038/sj.hdy.6800861
2. Protas ME, Patel NH. Evolution of Coloration Patterns. *Annu Rev Cell Dev Biol.* 2008; 24: 425–446. doi:10.1146/annurev.cellbio.24.110707.175302
3. Andersson L. Mutations in Domestic Animals Disrupting or Creating Pigmentation Patterns. *Front Ecol Evol.* 2020; 8. doi:10.3389/fevo.2020.00116
4. Silvers WK. The Coat Colors of Mice. The Coat Colors of Mice. New York, NY: *Springer New York*; 1979. doi:10.1007/978-1-4612-6164-3
5. Letko A, Ammann B, Jagannathan V, Henkel J, Leuthard F, Schelling C, *et al.* A deletion spanning the promoter and first exon of the hair cycle- specific ASIP transcript isoform in black and tan rabbits. *Anim Genet.* 2020; 51: 137–140. doi:10.1111/age.12881
6. Henkel J, Lafayette C, Brooks SA, Martin K, Patterson- Rosa L, Cook D, *et al.* Whole-genome sequencing reveals a large deletion in the *MITF* gene in horses with white spotted coat colour and increased risk of deafness. *Anim Genet.* 2019; 50: 172–174. doi:10.1111/age.12762
7. Hug P, Jude R, Henkel J, Jagannathan V, Leeb T. A novel KIT deletion variant in a German Riding Pony with white- spotting coat colour phenotype. *Anim Genet*. 2019; 50: 761–763. doi:10.1111/age.12840
8. Cichorek M, Wachulska M, Stasiewicz A, Tymińska A. Skin melanocytes: biology and development. *Adv Dermatology Allergol.* 2013; 1: 30–41. doi:10.5114/pdia.2013.33376
9. Wikipedia. Melanocyte. [cited 17 Feb 2021]. Available: https://en.wikipedia.org/wiki/Melanocyte
10. Mort RL, Jackson IJ, Patton EE. The melanocyte lineage in development and disease. *Development*. 2015; 142: 620–632. doi:10.1242/dev.106567
11. Cramer SF. The origin of epidermal melanocytes. Implications for the histogenesis of nevi and melanomas. *Arch Pathol Lab Med.* 1991; 115: 115–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/1992974
12. Thomas AJ, Erickson CA. The making of a melanocyte: The specification of melanoblasts from the neural crest. *Pigment Cell and Melanoma Research.* 2008; 21(6): 598-610. doi:10.1111/j.1755-148X.2008.00506.x
13. Shyamala K, Yanduri S, Girish H, Murgod S. Neural crest: The fourth germ layer. *J Oral Maxillofac Pathol*. 2015; 19: 221. doi:10.4103/0973-029X.164536
14. Rothstein M, Bhattacharya D, Simoes-Costa M. The molecular basis of neural crest axial identity. *Dev Biol*. 2018;444: S170–S180. doi:10.1016/j.ydbio.2018.07.026

15. Vandamme N, Berx G. From neural crest cells to melanocytes: cellular plasticity during development and beyond. *Cell Mol Life Sci.* 2019; 76: 1919–1934. doi:10.1007/s00018-019-03049-w
16. Aybar MJ, Mayor R. Early induction of neural crest cells: lessons learned from frog, fish and chick. *Curr Opin Genet Dev.* 2002; 12: 452–458. doi:10.1016/S0959-437X(02)00325-8
17. Anderson DJ. Cellular and molecular biology of neural crest cell lineage determination. *Trends Genet.* 1997; 13: 276–280. doi:10.1016/S0168-9525(97)01187-6
18. Greenhill ER, Rocco A, Vibert L, Nikaido M, Kelsh RN. An iterative genetic and dynamical modelling approach identifies novel features of the gene regulatory network underlying melanocyte development. *PLoS Genet*. 2011; 7: e1002265. doi:10.1371/journal.pgen.1002265
19. Green SA, Simoes-Costa M, Bronner ME. Evolution of vertebrates as viewed from the crest. *Nature*. 2015; 520: 474–482. doi:10.1038/nature14436
20. Hawkins K, Russell J, Murray A, Jurek P, Beutler B. Record for Widget. Dallas, TX; 2016 [cited 15 Feb 2021] p. 16. Available: https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=1519
21. Reissmann M, Ludwig A. Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Semin Cell Dev Biol*. 2013; 24: 576–586. doi:10.1016/j.semcdb.2013.03.014
22. Smith SD. Tietz syndrome (hypopigmentation/deafness) caused by mutation of MITF. *J Med Genet*. 2000; 37: 446–448. doi:10.1136/jmg.37.6.446
23. Metallinos DL, Bowling AT, Rine J. A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mamm Genome*. 1998; 9: 426–31. doi:10.1007/s003359900790
24. Besmer P, Manova K, Duttlinger R, Huang EJ, Packer A, Gyssler C, *et al.* The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis. *Dev Suppl*. 1993; 125–37. Available: http://www.ncbi.nlm.nih.gov/pubmed/7519481
25. Fleischman RA. From white spots to stem cells: the role of the Kit receptor in mammalian development. *Trends Genet*. 1993; 9: 285–290. doi:10.1016/0168-9525(93)90015-A
26. WAARDENBURG PJ. A new syndrome combining developmental anomalies of the eyelids, eyebrows and nose root with pigmentary defects of the iris and head hair and with congenital deafness. *Am J Hum Genet*. 1951; 3: 195–253. Available: http://www.ncbi.nlm.nih.gov/pubmed/14902764
27. Pingault V, Ente D, Dastot-Le Moal F, Goossens M, Marlin S, Bondurand N. Review

and update of mutations causing Waardenburg syndrome. *Hum Mutat*. 2010; 31: 391–406. doi:10.1002/humu.21211

28. Hachiya A, Kobayashi A, Ohuchi A, Takema Y, Imokawa G. The paracrine role of stem cell factor/c-kit signaling in the activation of human melanocytes in ultraviolet-B-induced pigmentation. *J Invest Dermatol*. 2001; 116: 578–86. doi:10.1046/j.1523-1747.2001.01290.x
29. Roskoski R. Signaling by Kit protein-tyrosine kinase—The stem cell factor receptor. *Biochem Biophys Res Commun*. 2005; 337: 1–13. doi:10.1016/j.bbrc.2005.08.055
30. Lennartsson J, Rönnstrand L. Stem Cell Factor Receptor/c-Kit: From Basic Science to Clinical Implications. *Physiol Rev.* 2012; 92: 1619–1649. doi:10.1152/physrev.00046.2011
31. Geissler EN, Ryan MA, Housman DE. The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell*. 1988; 55: 185–192. doi:10.1016/0092-8674(88)90020-7
32. Giebel LB, Spritz RA. Mutation of the KIT (mast/stem cell growth factor receptor) protooncogene in human piebaldism. *Proc Natl Acad Sci.* 1991; 88: 8696–8699. doi:10.1073/pnas.88.19.8696
33. Fleischman RA, Saltman DL, Stastny V, Zneimer S. Deletion of the c-kit protooncogene in the human developmental defect piebald trait. *Proc Natl Acad Sci*. 1991; 88: 10885–10889. doi:10.1073/pnas.88.23.10885
34. Haase B, Brooks SA, Tozaki T, Burger D, Poncet P-A, Rieder S, *et al.* Seven novel KIT mutations in horses with white coat colour phenotypes. *Anim Genet*. 2009; 40: 623–9. doi:10.1111/j.1365-2052.2009.01893.x
35. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, *et al.* Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci.* 2012; 109: 19529–19536. doi:10.1073/pnas.1217149109
36. Oiso N, Fukai K, Kawada A, Suzuki T. Piebaldism. *J Dermatol*. 2013; 40: 330–335. doi:10.1111/j.1346-8138.2012.01583.x
37. Aoki H, Tomita H, Hara A, Kunisada T. Conditional Deletion of Kit in Melanocytes: White Spotting Phenotype Is Cell Autonomous. *J Invest Dermatol*. 2015; 135: 1829–1838. doi:10.1038/jid.2015.83
38. Webb ML, Chao CC, Rizzo M, Shapiro RA, Neubauer M, Liu ECK, *et al.* Cloning and expression of an endothelin receptor subtype B from human prostate that mediates contraction. *Mol Pharmacol*. 1995; 47: 730–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/7536888
39. Mccallion AS, Chakravarti A. *EDNRB/EDN3* and Hirschsprung Disease Type II. *Pigment*

Cell Res. 2001; 14: 161–169. doi:10.1034/j.1600-0749.2001.140305.x

40. Eberle J, Weitmann S, Thieck O, Pech H, Orfanos CE, Paul M. Downregulation of Endothelin B Receptor in Human Melanoma Cell Lines Parallel to Differentiation Genes. *J Invest Dermatol*. 1999; 112: 925–932. doi:10.1046/j.1523-1747.1999.00598.x
41. Lahav R. Endothelin receptor B is required for the expansion of melanocyte precursors and malignant melanoma. *Int J Dev Biol*. 2005; 49: 173–180. doi:10.1387/ijdb.041951rl
42. Van Raamsdonk CD, Fitch KR, Fuchs H, de Angelis MH, Barsh GS. Effects of G-protein mutations on skin color. *Nat Genet*. 2004; 36: 961–968. doi:10.1038/ng1412
43. Stanchina L, Baral V, Robert F, Pingault V, Lemort N, Pachnis V, *et al.* Interactions between Sox10, Edn3 and Ednrb during enteric nervous system and melanocyte development. *Dev Biol.* 2006; 295: 232–249. doi:10.1016/j.ydbio.2006.03.031
44. Pla P, Larue L. Involvement of endothelin receptors in normal and pathological development of neural crest cells. *Int J Dev Biol.* 2003; 47: 315–25. doi:10.1387/ijdb.12895026
45. Garcia RJ, Ittah A, Mirabal S, Figueroa J, Lopez L, Glick AB, *et al.* Endothelin 3 Induces Skin Pigmentation in a Keratin-Driven Inducible Mouse Model. *J Invest Dermatol.* 2008; 128: 131–142. doi:10.1038/sj.jid.5700948
46. Hemesath TJ, Steingrimsson E, McGill G, Hansen MJ, Vaught J, Hodgkinson CA, *et al.* microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. *Genes Dev.* 1994; 8: 2770–2780. doi:10.1101/gad.8.22.2770
47. Shibahara S, Takeda K, Yasumoto K, Udono T, Watanabe K, Saito H, *et al.* Microphthalmia-Associated Transcription Factor (MITF): Multiplicity in Structure, Function, and Regulation. *J Investig Dermatology* Symp Proc. 2001; 6: 99–104. doi:10.1046/j.0022-202x.2001.00010.x
48. Steingrímsson E, Copeland NG, Jenkins NA. Melanocytes and the microphthalmia transcription factor network. *Annu Rev Genet.* 2004; 38: 365–411. doi:10.1146/annurev.genet.38.072902.092717
49. Levy C, Khaled M, Fisher DE. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med.* 2006; 12: 406–414. doi:10.1016/j.molmed.2006.07.008
50. Goding CR, Arnheiter H. MITF—the first 25 years. *Genes Dev.* 2019; 33: 983–1007. doi:10.1101/gad.324657.119
51. Hauswirth R, Haase B, Blatter M, Brooks SA, Burger D, Drögemüller C, *et al.* Mutations in *MITF* and *PAX3* Cause “Splashed White” and Other White Spotting Phenotypes in Horses. Barsh GS, editor. *PLoS Genet.* 2012; 8: e1002653. doi:10.1371/journal.pgen.1002653

52. Baranowska Körberg I, Sundström E, Meadows JRS, Rosengren Pielberg G, Gustafson U, Hedhammar Å, *et al.* A Simple Repeat Polymorphism in the *MITF-M* Promoter Is a Key Regulator of White Spotting in Dogs. *Murphy WJ*, editor. PLoS One. 2014; 9: e104363. doi:10.1371/journal.pone.0104363
53. Magdesian KG, Tanaka J, Bellone RR. A De Novo MITF Deletion Explains a Novel Splashed White Phenotype in an American Paint Horse. Bailey E, editor. *J Hered*. 2020; 111: 287–293. doi:10.1093/jhered/esaa009
54. Steingrímsson E, Moore KJ, Lamoreux ML, Ferré-D'Amaré AR, Burley SK, Zimring DCS, *et al.* Molecular basis of mouse microphthalmia (mi) mutations helps explain their developmental and phenotypic consequences. *Nat Genet.* 1994; 8: 256–63. doi:10.1038/ng1194-256
55. Yasumoto K, Takeda K, Saito H, Watanabe K, Takahashi K, Shibahara S. Microphthalmia-associated transcription factor interacts with LEF-1, a mediator of Wnt signaling. *EMBO J*. 2002; 21: 2703–14. doi:10.1093/emboj/21.11.2703
56. Hallsson JH, Haflidadóttir BS, Stivers C, Odenwald W, Arnheiter H, Pignoni F, *et al.* The Basic Helix-Loop-Helix Leucine Zipper Transcription Factor Mitf Is Conserved in Drosophila and Functions in Eye Development. *Genetics*. 2004; 167: 233–241. doi:10.1534/genetics.167.1.233
57. Osawa M. Melanocyte stem cells. *StemBook*. 2008. doi:10.3824/stembook.1.46.1
58. Harris ML, Baxter LL, Loftus SK, Pavan WJ. Sox proteins in melanocyte development and melanoma. *Pigment Cell Melanoma Res.* 2010; 23: 496–513. doi:10.1111/j.1755-148X.2010.00711.x
59. Haldin CE, LaBonne C. SoxE factors as multifunctional neural crest regulatory factors. *Int J Biochem Cell Biol.* 2010; 42: 441–4. doi:10.1016/j.biocel.2009.11.014
60. Seberg HE, Van Otterloo E, Cornell RA. Beyond MITF: Multiple transcription factors directly regulate the cellular phenotype in melanocytes and melanoma. *Pigment Cell Melanoma Res.* 2017; 30: 454–466. doi:10.1111/pcmr.12611
61. Herbarth B, Pingault V, Bondurand N, Kuhlbrodt K, Hermans-Borgmeyer I, Puliti A, *et al.* Mutation of the Sry-related Sox10 gene in Dominant megacolon, a mouse model for human Hirschsprung disease. *Proc Natl Acad Sci.* 1998; 95: 5161–5165. doi:10.1073/pnas.95.9.5161
62. Bondurand N, Dastot-Le Moal F, Stanchina L, Collot N, Baral V, Marlin S, *et al.* Deletions at the *SOX10* Gene Locus Cause Waardenburg Syndrome Types 2 and 4. *Am J Hum Genet.* 2007; 81: 1169–1185. doi:10.1086/522090
63. Tremblay P, Gruss P. Pax: Genes for mice and men. *Pharmacol Ther.* 1994; 61: 205–226. doi:10.1016/0163-7258(94)90063-9

64. Kubic JD, Young KP, Plummer RS, Ludvik AE, Lang D. Pigmentation PAX-ways: the role of Pax3 in melanogenesis, melanocyte stem cell maintenance, and disease. *Pigment Cell Melanoma Res.* 2008; 21: 627–45. doi:10.1111/j.1755-148X.2008.00514.x
65. Boudjadi S, Chatterjee B, Sun W, Vemu P, Barr FG. The expression and function of PAX3 in development and disease. *Gene*. 2018; 666: 145–157. doi:10.1016/j.gene.2018.04.087
66. Watanabe A, Takeda K, Ploplis B, Tachibana M. Epistatic relationship between Waardenburg Syndrome genes *MITF* and *PAX3*. *Nat Genet.* 1998; 18: 283–286. doi:10.1038/ng0398-283
67. Aybar MJ. Snail precedes Slug in the genetic cascade required for the specification and migration of the Xenopus neural crest. *Development*. 2003; 130: 483–494. doi:10.1242/dev.00238
68. Nieto MA. The snail superfamily of zinc-finger transcription factors. *Nat Rev Mol Cell Biol.* 2002; 3: 155–166. doi:10.1038/nrm757
69. Nieto M, Sargent M, Wilkinson D, Cooke J. Control of cell behavior during vertebrate development by *Slug*, a zinc finger gene. *Science* (80-). 1994; 264: 835–839. doi:10.1126/science.7513443
70. Tríbulo C, Aybar MJ, Sánchez SS, Mayor R. A balance between the anti-apoptotic activity of Slug and the apoptotic activity of msx1 is required for the proper development of the neural crest. *Dev Biol.* 2004; 275: 325–342. doi:10.1016/j.ydbio.2004.07.041
71. Cobaleda C, Pérez-Caro M, Vicente-Dueñas C, Sánchez-García I. Function of the Zinc-Finger Transcription Factor SNAI2 in Cancer and Development. *Annu Rev Genet.* 2007; 41: 41–61. doi:10.1146/annurev.genet.41.110306.130146
72. Pérez-Losada J, Sánchez-Martín M, Rodríguez-García A, Sánchez ML, Orfao A, Flores T, *et al.* Zinc-finger transcription factor Slug contributes to the function of the stem cell factor c-kit signaling pathway. *Blood*. 2002; 100: 1274–1286.
73. Sun G, Liang X, Qin K, Qin Y, Shi X, Cong P, *et al.* Functional Analysis of *KIT* Gene Structural Mutations Causing the Porcine Dominant White Phenotype Using Genome Edited Mouse Models. *Front Genet.* 2020;11: 138. doi:10.3389/fgene.2020.00138
74. Saldana-Caboverde A, Kos L. Roles of endothelin signaling in melanocyte development and melanoma. *Pigment Cell Melanoma Res.* 2010; 23: 160–170. doi:10.1111/j.1755-148X.2010.00678.x
75. Hallsson JH, Favor J, Hodgkinson C, Glaser T, Lamoreux ML, Magnúsdóttir R, *et al.* Genomic, transcriptional and mutational analysis of the mouse microphthalmia locus. *Genetics*. 2000; 155: 291–300. Available: http://www.ncbi.nlm.nih.gov/pubmed/10790403

76. Siggs O, Smart NG, Bruce B. Record for Dalmatian. Dallas, TX; 2016 [cited 15 Feb 2021] p. 17. Available: https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=265
77. Ohno T, Maegawa T, Katoh H, Miyasaka Y, Suzuki M, Kobayashi M, *et al.* A new missense mutation in the paired domain of the mouse *Pax3* gene. *Exp Anim.* 2017; 66: 245–250. doi:10.1538/expanim.17-0013
78. Sánchez-Martín M, Pérez-Losada J, Rodríguez-García A, González-Sánchez B, Korf BR, Kuster W, *et al.* Deletion of the SLUG (*SNAI2*) gene results in human piebaldism. *Am J Med Genet Part A*. 2003; 122A: 125–132. doi:10.1002/ajmg.a.20345
79. The Jackson Laboratory. Strain #000689. [cited 4 Mar 2021]. Available: https://www.jax.org/strain/000689
80. Percival GH, Stewart CP. Melanogenesis: A Review. *Edinb Med J.* 1930; 37: 497–523. Available: http://www.ncbi.nlm.nih.gov/pubmed/29639202
81. Hearing VJ. Determination of Melanin Synthetic Pathways. *J Invest Dermatol.* 2011; 131: E8–E11. doi:10.1038/skinbio.2011.4
82. Bultman SJ, Michaud EJ, Woychik RP. Molecular characterization of the mouse agouti locus. *Cell*. 1992; 71: 1195–1204. doi:10.1016/S0092-8674(05)80067-4
83. Barsh GS. The genetics of pigmentation: from fancy genes to complex traits. *Trends Genet.* 1996; 12: 299–305. doi:10.1016/0168-9525(96)10031-7
84. Vrieling H, Duhl DMJ, Millar SE, Miller KA, Barsh GS. Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proc Natl Acad Sci.* 1994; 91: 5667–5671. doi:10.1073/pnas.91.12.5667
85. Millar SE, Miller MW, Stevens ME, Barsh GS. Expression and transgenic studies of the mouse agouti gene provide insight into the mechanisms by which mammalian coat color patterns are generated. *Development*. 1995; 121: 3223–32. Available: http://www.ncbi.nlm.nih.gov/pubmed/7588057
86. Le Pape E, Wakamatsu K, Ito S, Wolber R, Hearing VJ. Regulation of eumelaninpheomelanin synthesis and visible pigmentation in melanocytes by ligands of the melanocortin 1 receptor. *Pigment Cell Melanoma Res.* 2008; 21: 477–486. doi:10.1111/j.1755-148X.2008.00479.x
87. Wolf Horrell EM, Boulanger MC, D'Orazio JA. Melanocortin 1 Receptor: Structure, Function, and Regulation. *Front Genet.* 2016; 7: 95. doi:10.3389/fgene.2016.00095
88. Kondo T, Hearing VJ. Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Rev Dermatol.* 2011; 6: 97–108. doi:10.1586/edm.10.70
89. Woolf CM. Common white facial markings in Arabian horses that are homozygous and heterozygous for alleles at the A and E loci. *J Hered.* 1992; 83: 73–7.

doi:10.1093/oxfordjournals.jhered.a111163

90. Woolf CM. Influence of stochastic events on the phenotypic variation of common white leg markings in the Arabian horse: implications for various genetic disorders in humans. *J Hered*. 1995; 86: 129–35. doi:10.1093/oxfordjournals.jhered.a111542
91. Haase B, Signer-Hasler H, Binns MM, Obexer-Ruff G, Hauswirth R, Bellone RR, *et al.* Accumulating mutations in series of haplotypes at the *KIT* and *MITF* loci are major determinants of white markings in Franches-Montagnes horses. *PLoS One*. 2013; 8: e75071. doi:10.1371/journal.pone.0075071
92. Summers CG. Albinism: Classification, Clinical Characteristics, and Recent Findings. *Optom Vis Sci*. 2009; 86: 659–662. doi:10.1097/OPX.0b013e3181a5254c
93. Marçon CR, Maia M. Albinism: epidemiology, genetics, cutaneous characterization, psychosocial factors. *An Bras Dermatol*. 2019; 94: 503–520. doi:10.1016/j.abd.2019.09.023
94. Schiaffino MV, Bassi MT, Galli L, Renieri A, Bruttini M, Nigris F De, *et al.* Analysis of the *OA1* gene reveals mutations in only one-third of patients with X-linked ocular albinism. *Hum Mol Genet*. 1995; 4: 2319–2325. doi:10.1093/hmg/4.12.2319
95. d'Addio M. Defective intracellular transport and processing of OA1 is a major cause of ocular albinism type 1. *Hum Mol Genet*. 2000; 9: 3011–3018. doi:10.1093/hmg/9.20.3011
96. Hulsman Hanna LL, Sanders JO, Riley DG, Abbey CA, Gill CA. Identification of a major locus interacting with MC1R and modifying black coat color in an F2 Nellore-Angus population. *Genet Sel Evol*. 2014; 46: 4. doi:10.1186/1297-9686-46-4
97. Gonçalves GL, Paixão-Côrtes VR, Freitas TRO. Molecular evolution of the melanocortin 1-receptor pigmentation gene in rodents. *Genet Mol Res*. 2013; 12: 3230–45. doi:10.4238/2013.February.28.24
98. Lu D, Willard D, Patel IR, Kadwell S, Overton L, Kost T, *et al.* Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. *Nature*. 1994; 371: 799–802. doi:10.1038/371799a0
99. Graham A, Wakamatsu K, Hunt G, Ito S, Thody AJ. Agouti protein inhibits the production of eumelanin and phaeomelanin in the presence and absence of alpha-melanocyte stimulating hormone. *Pigment cell Res*. 1997; 10: 298–303. doi:10.1111/j.1600-0749.1997.tb00689.x
100. Suzuki I, Tada A, Ollmann MM, Barsh GS, Im S, Lynn Lamoreux M, *et al.* Agouti Signaling Protein Inhibits Melanogenesis and the Response of Human Melanocytes to α-Melanotropin. *J Invest Dermatol*. 1997; 108: 838–842. doi:10.1111/1523-1747.ep12292572

101. Spritz RA. Molecular genetics of oculocutaneous albinism. *Hum Mol Genet*. 1994; 3 Spec No: 1469–75. doi:10.1093/hmg/3.suppl_1.1469
102. Bouchard B, Marmol V, Jackson IJ, Cherif D, Dubertret L. Molecular characterization of a human tyrosinase-related-protein-2 cDNA. Patterns of expression in melanocytic cells. *Eur J Biochem*. 1994; 219: 127–134. doi:10.1111/j.1432-1033.1994.tb19922.x
103. Pennamen P, Tingaud-Sequeira A, Gazova I, Keighren M, McKie L, Marlin S, *et al.* Dopachrome tautomerase variants in patients with oculocutaneous albinism. *Genet Med.* 2020. doi:10.1038/s41436-020-00997-8
104. Steel KP, Davidson DR, Jackson IJ. TRP-2/DT, a new early melanoblast marker, shows that steel growth factor (c-kit ligand) is a survival factor. *Development*. 1992; 115: 1111–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/1280558
105. Jiménez-Cervantes C, Solano F, Kobayashi T, Urabe K, Hearing VJ, Lozano JA, *et al.* A new enzymatic function in the melanogenic pathway. The 5,6-dihydroxyindole-2-carboxylic acid oxidase activity of tyrosinase-related protein-1 (TRP1). *J Biol Chem*. 1994; 269: 17993–18000. doi:10.1016/S0021-9258(17)32408-0
106. Kobayashi T, Urabe K, Winder A, Jiménez-Cervantes C, Imokawa G, Brewington T, *et al.* Tyrosinase related protein 1 (TRP1) functions as a DHICA oxidase in melanin biosynthesis. *EMBO J.* 1994; 13: 5818–5825. doi:10.1002/j.1460-2075.1994.tb06925.x
107. Boissy RE, Zhao H, Oetting WS, Austin LM, Wildenberg SC, Boissy YL, *et al.* Mutation in and lack of expression of tyrosinase-related protein-1 (TRP-1) in melanocytes from an individual with brown oculocutaneous albinism: a new subtype of albinism classified as “OCA3”. *Am J Hum Genet.* 1996; 58: 1145–56. Available: http://www.ncbi.nlm.nih.gov/pubmed/8651291
108. Manga P, Kromberg JGR, Box NF, Sturm RA, Jenkins T, Ramsay M. Rufous Oculocutaneous Albinism in Southern African Blacks Is Caused by Mutations in the *TYRP1* Gene. *Am J Hum Genet*. 1997; 61: 1095–1101. doi:10.1086/301603
109. Zdarsky E, Favor J, Jackson IJ. The molecular basis of brown, an old mouse mutation, and of an induced revertant to wild type. *Genetics*. 1990; 126: 443–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/2245916
110. Becker D, Otto M, Ammann P, Keller I, Drögemüller C, Leeb T. The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the *TYRP1* locus on chromosome 8. *Anim Genet*. 2015; 46: 50–54. doi:10.1111/age.12240
111. Dietrich J, Menzi F, Ammann P, Drögemüller C, Leeb T. A breeding experiment confirms the dominant mode of inheritance of the brown coat colour associated with the 496 Asp TYRP1 allele in goats. *Anim Genet.* 2015; 46: 587–588. doi:10.1111/age.12320
112. Paris JM, Letko A, Häfliger IM, Ammann P, Flury C, Drögemüller C. Identification of two

TYRP1 loss- of- function alleles in Valais Red sheep. *Anim Genet*. 2019; 50: 778–782. doi:10.1111/age.12863

113. Price EO. Behavioral Aspects of Animal Domestication. *Q Rev Biol*. 1984; 59: 1–32. doi:10.1086/413673
114. Mignon-Grasteau S, Boissy A, Bouix J, Faure J-M, Fisher AD, Hinch GN, *et al.* Genetics of adaptation and domestication in livestock. *Livest Prod Sci*. 2005; 93: 3–14. doi:10.1016/j.livprodsci.2004.11.001
115. Darwin C. On the Origin of the Species. *London: Jo*. Darwin. 1859.
116. Nielsen R. Molecular Signatures of Natural Selection. *Annu Rev Genet*. 2005; 39: 197–218. doi:10.1146/annurev.genet.39.073003.112420
117. Vitti JJ, Grossman SR, Sabeti PC. Detecting Natural Selection in Genomic Data. *Annu Rev Genet*. 2013; 47: 97–120. doi:10.1146/annurev-genet-111212-133526
118. Raeside JI. The Behaviour of Domestic Animals. *Can Vet J*. 1963; 4: 120. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1695275/
119. Zeder MA, Emshwiller E, Smith BD, Bradley DG. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet*. 2006; 22: 139–55. doi:10.1016/j.tig.2006.01.007
120. Zeder MA. Pathways to Animal Domestication. In: Gepts P, Famula TR, Bettinger RL, Brush SB, Damania AB, McGuire PE, *et al.*, editors. Biodiversity in Agriculture. Cambridge: *Cambridge University Press*; 2012. pp. 227–259. doi:10.1017/CBO9781139019514.013
121. Andersson L. Molecular consequences of animal breeding. *Curr Opin Genet Dev*. 2013; 23: 295–301. doi:10.1016/j.gde.2013.02.014
122. Andersson L. Domestic animals as models for biomedical research. *Ups J Med Sci*. 2016; 121: 1–11. doi:10.3109/03009734.2015.1091522
123. Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L. Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. Barsh GS, editor. *PLoS Genet*. 2009; 5: e1000341. doi:10.1371/journal.pgen.1000341
124. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012; 491: 393–8. doi:10.1038/nature11622
125. Padmanabhan R, Padmanabhan R, Wu R. Nucleotide sequence analysis of DNA. *Biochem Biophys Res Commun.* 1972; 48: 1295–1302. doi:10.1016/0006-291X(72)90852-2
126. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci.* 1977; 74: 5463–5467. doi:10.1073/pnas.74.12.5463

127. Saiki R, Scharf S, Faloona F, Mullis K, Horn G, Erlich H, *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* (80-). 1985; 230: 1350–1354. doi:10.1126/science.2999980
128. Collins FS. Positional cloning moves from perditional to traditional. *Nat Genet.* 1995; 9: 347–350. doi:10.1038/ng0495-347
129. Vink JM, Boomsma D. Gene finding strategies. *Biol Psychol*. 2002; 61: 53–71. doi:10.1016/S0301-0511(02)00052-2
130. Vahidnezhad H, Youssefian L, Jazayeri A, Uitto J. Research Techniques Made Simple: Genome-Wide Homozygosity/Autozygosity Mapping Is a Powerful Tool for Identifying Candidate Genes in Autosomal Recessive Genetic Diseases. *J Invest Dermatol.* 2018; 138: 1893–1900. doi:10.1016/j.jid.2018.06.170
131. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SPA. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci*. 1994; 91: 5022–5026. doi:10.1073/pnas.91.11.5022
132. Djakow J, Kramná L, Dušátková L, Uhlík J, Pursiheimo J-P, Svobodová T, *et al.* An effective combination of sanger and next generation sequencing in diagnostics of primary ciliary dyskinesia. *Pediatr Pulmonol.* 2016; 51: 498–509. doi:10.1002/ppul.23261
133. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell*. 2013; 155: 27–38. doi:10.1016/j.cell.2013.09.006
134. Gauthier J, Vincent AT, Charette SJ, Derome N. A brief history of bioinformatics. *Brief Bioinform.* 2019; 20: 1981–1996. doi:10.1093/bib/bby063
135. McGinn S, Gut IG. DNA sequencing – spanning the generations. *N Biotechnol.* 2013; 30: 366–372. doi:10.1016/j.nbt.2012.11.012
136. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016; 107: 1–8. doi:10.1016/j.ygeno.2015.11.003
137. Suzuki S, Ono N, Furusawa C, Ying B-W, Yomo T. Comparison of Sequence Reads Obtained from Three Next-Generation Sequencing Platforms. Bereswill S, editor. PLoS One. 2011; 6: e19534. doi:10.1371/journal.pone.0019534
138. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011; 475: 348–352. doi:10.1038/nature10242
139. Jagannathan V, Drögemüller C, Leeb T, Aguirre G, André C, Bannasch D, *et al.* A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves*. Anim Genet*. 2019; 50: 695–704. doi:10.1111/age.12834

140. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, *et al.* A global reference for human genetic variation. *Nature*. 2015; 526: 68–74. doi:10.1038/nature15393
141. Hug P, Anderegg L, Kehl A, Jagannathan V, Leeb T. AKNA Frameshift Variant in Three Dogs with Recurrent Inflammatory Pulmonary Disease. *Genes (Basel).* 2019; 10: 567. doi:10.3390/genes10080567
142. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet.* 2014; 30: 418–26. doi:10.1016/j.tig.2014.07.001
143. Haque F, Li J, Wu H-C, Liang X-J, Guo P. Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA. *Nano Today*. 2013; 8: 56–74. doi:10.1016/j.nantod.2012.12.008
144. Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*. 2009; 4: 265–270. doi:10.1038/nnano.2009.12
145. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet.* 2010; 19: R227–R240. doi:10.1093/hmg/ddq416
146. Feingold EA, Good PJ, Guyer MS, Kamholz S, Liefer L, Wetterstrand K, *et al.* The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* (80-). 2004; 306: 636–640. doi:10.1126/science.1105136
147. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489: 57–74. doi:10.1038/nature11247
148. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20: 1297–1303. doi:10.1101/gr.107524.110
149. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet.* 2011; 12: 363–376. doi:10.1038/nrg2958
150. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*. 2009; 6: 677–681. doi:10.1038/nmeth.1363
151. Fiegler H, Redon R, Andrews D, Scott C, Andrews R, Carder C, *et al.* Accurate and reliable high-throughput detection of copy number variation in the human genome. *Genome Res*. 2006; 16: 1566–1574. doi:10.1101/gr.5630906
152. Huddleston J, Chaisson MJP, Steinberg KM, Warren W, Hoekzema K, Gordon D, *et al.* Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res*. 2017; 27: 677–685. doi:10.1101/gr.214007.116

153. Fang L, Hu J, Wang D, Wang K. NextSV: a meta-caller for structural variants from low-coverage long-read sequencing data. *BMC Bioinformatics*. 2018; 19: 180. doi:10.1186/s12859-018-2207-1

154. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. *Nat Rev Genet*. 2019/11/15. 2020; 21: 171–189. doi:10.1038/s41576-019-0180-9

155. Gunnarsson U, Kerje S, Bed'hom B, Sahlqvist A-S, Ekwall O, Tixier-Boichard M, *et al.* The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. *Pigment Cell Melanoma Res*. 2011; 24: 268–74. doi:10.1111/j.1755-148X.2011.00825.x

156. Brooks SA, Lear TL, Adelson DL, Bailey E. A chromosome inversion near the *KIT* gene and the Tobiano spotting pattern in horses. *Cytogenet Genome Res*. 2007; 119: 225–230. doi:10.1159/000112065

157. Grahofer A, Letko A, Häfliger IM, Jagannathan V, Ducos A, Richard O, *et al.* Chromosomal imbalance in pigs showing a syndromic form of cleft palate. *BMC Genomics.* 2019; 20: 349. doi:10.1186/s12864-019-5711-4

158. Simon R, Lischer HEL, Pieńkowska- Schelling A, Keller I, Häfliger IM, Letko A, *et al.* New genomic features of the polled intersex syndrome variant in goats unraveled by long- read whole- genome sequencing. *Anim Genet*. 2020; 51: 439–448. doi:10.1111/age.12918

159. Knight J. Hypothesis-free research. *Trends Genet*. 2000; 16: 25. doi:10.1016/S0168-9525(00)00104-9

160. Teare MD, Santibanez Koref MF. Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. *Brief Funct Genomics*. 2014; 13: 378–383. doi:10.1093/bfgp/elu024

161. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. *Lancet*. 2005; 366: 941–951. doi:10.1016/S0140-6736(05)67322-9

162. Teo YY. Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr Opin Lipidol.* 2008; 19: 133–143. doi:10.1097/MOL.0b013e3282f5dd77

163. Jensen JD, Foll M, Bernatchez L. The past, present and future of genomic scans for selection. *Mol Ecol.* 2016; 25: 1–4. doi:10.1111/mec.13493

164. Otto SP. Detecting the form of selection from DNA sequence data. *Trends Genet.* 2000; 16: 526–529. doi:10.1016/S0168-9525(00)02141-7

165. Saravanan KA, Panigrahi M, Kumar H, Bhushan B, Dutt T, Mishra BP. Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livest Sci.* 2020; 241: 104257. doi:10.1016/j.livsci.2020.104257

166. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*. 2010; 464: 587–591. doi:10.1038/nature08832
167. Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* (80-). 2014; 345: 1074–1079. doi:10.1126/science.1253714
168. Martinez Barrio A, Lamichhaney S, Fan G, Rafati N, Pettersson M, Zhang H, *et al.* The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *Elife*. 2016; 5. doi:10.7554/eLife.12081
169. WRIGHT S. The genetical structure of populations. *Ann Eugen*. 1951; 15: 323–54. doi:10.1111/j.1469-1809.1949.tb02451.x
170. Bourneuf E, Otz P, Pausch H, Jagannathan V, Michot P, Grohs C, *et al.* Rapid Discovery of De Novo Deleterious Mutations in Cattle Enhances the Value of Livestock as Model Species. *Sci Rep.* 2017; 7: 11466. doi:10.1038/s41598-017-11523-3
171. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* (Austin). 2012; 6: 80–92. doi:10.4161/fly.19695
172. Durkin K, Coppieters W, Drögemüller C, Ahariz N, Cambisano N, Druet T, *et al.* Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*. 2012; 482: 81–84. doi:10.1038/nature10757
173. Küttel L, Letko A, Häfliger IM, Signer- Hasler H, Joller S, Hirsbrunner G, *et al.* A complex structural variant at the *KIT* locus in cattle with the Pinzgauer spotting pattern. *Anim Genet.* 2019; 50: 423–429. doi:10.1111/age.12821
174. Artesi M, Tamma N, Deckers M, Karim L, Coppieters W, Van den Broeke A, *et al.* Colour- sidedness in Gloucester cattle is associated with a complex structural variant impacting regulatory elements downstream of *KIT*. *Anim Genet.* 2020; 51: 461–465. doi:10.1111/age.12932
175. Borter J. Die Walliser Schwarzhalsziege. *Swiss Library Service Platform.* 1975.
176. Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, *et al.* Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Anim Genet.* 2016; 47: 727–739. doi:10.1111/age.12476
177. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet Sel Evol*. 2018; 50: 57. doi:10.1186/s12711-018-0421-y
178. Montes DE, Braz CU, Ribeiro AMF, Cavani L, Barbero MMD, Albuquerque LG, *et al.* Selection signatures in candidate genes and QTL for reproductive traits in Nellore

heifers. *Anim Reprod Sci.* 2019; 207: 1–8. doi:10.1016/j.anireprosci.2019.06.004

179. Signer- Hasler H, Burren A, Ammann P, Drögemüller C, Flury C. Runs of homozygosity and signatures of selection: a comparison among eight local Swiss sheep breeds. *Anim Genet.* 2019; 50: 512–525. doi:10.1111/age.12828
180. Friedrich J, Wiener P. Selection signatures for high- altitude adaptation in ruminants. *Anim Genet.* 2020; 51: 157–165. doi:10.1111/age.12900
181. Saif R, Henkel J, Jagannathan V, Drögemüller C, Flury C, Leeb T. The *LCORL* Locus Is under Selection in Large-Sized Pakistani Goat Breeds. *Genes (Basel).* 2020; 11: 168. doi:10.3390/genes11020168
182. Marian AJ. Causality in genetics: the gradient of genetic effects and back to Koch's postulates of causality. *Circ Res.* 2014; 114: e18-21. doi:10.1161/CIRCRESAHA.114.302904
183. Awasthi Mishra N, Drögemüller C, Jagannathan V, Keller I, Wüthrich D, Bruggmann R, *et al.* A structural variant in the 5'-flanking region of the *TWIST2* gene affects melanocyte development in belted cattle. Tang H, editor. *PLoS One.* 2017; 12: e0180170. doi:10.1371/journal.pone.0180170
184. Holl H, Brooks S, Bailey E. *De novo* mutation of *KIT* discovered as a result of a non-hereditary white coat colour pattern. *Anim Genet*. 2010; 41: 196–198. doi:10.1111/j.1365-2052.2010.02135.x
185. Hoban R, Castle K, Hamilton N, Haase B. Novel KIT variants for dominant white in the Australian horse population. *Anim Genet*. 2018; 49: 99–100. doi:10.1111/age.12627
186. Haase B, Brooks SA, Schlumbaum A, Azor PJ, Bailey E, Alaeddine F, *et al.* Allelic Heterogeneity at the Equine *KIT* Locus in Dominant White (W) Horses. McVean G, editor. *PLoS Genet*. 2007; 3: e195. doi:10.1007/s00335-005-2472-y
187. Holl HM, Brooks SA, Carpenter ML, Bustamante CD, Lafayette C. A novel splice mutation within equine *KIT* and the W15 allele in the homozygous state lead to all white coat color phenotypes. *Anim Genet*. 2017; 48: 497–498. doi:10.1111/age.12554
188. Hauswirth R, Jude R, Haase B, Bellone RR, Archer S, Holl H, *et al.* Novel variants in the *KIT* and *PAX3* genes in horses with white-spotted coat colour phenotypes. *Anim Genet.* 2013; 44: 763–765. doi:10.1111/age.12057
189. Haase B, Rieder S, Tozaki T, Hasegawa T, Penedo MCT, Jude R, *et al.* Five novel *KIT* mutations in horses with white coat colour phenotypes. *Anim Genet.* 2011; 42: 337–339. doi:10.1111/j.1365-2052.2011.02173.x
190. Dürig N, Jude R, Holl H, Brooks SA, Lafayette C, Jagannathan V, *et al.* Whole genome sequencing reveals a novel deletion variant in the *KIT* gene in horses with white spotted coat colour phenotypes. *Anim Genet.* 2017; 48: 483–485. doi:10.1111/age.12556

191. Martin K, Patterson Rosa L, Vierra M, Foster G, Brooks SA, Lafayette C. *De novo* mutation of KIT causes extensive coat white patterning in a family of Berber horses. *Anim Genet.* 2021; 52: 135–137. doi:10.1111/age.13017
192. Capomaccio S, Milanesi M, Nocelli C, Giontella A, Verini-Supplizi A, Branca M, *et al.* Splicing site disruption in the *KIT* gene as strong candidate for white dominant phenotype in an Italian Trotter. *Anim Genet.* 2017; 48: 727–728. doi:10.1111/age.12590
193. Brooks SA, Bailey E. Exon skipping in the *KIT* gene causes a Sabino spotting pattern in horses. *Mamm Genome*. 2005; 16: 893–902. doi:10.1007/s00335-005-2472-y
194. Haase B, Jagannathan V, Rieder S, Leeb T. A novel *KIT* variant in an Icelandic horse with white-spotted coat colour. *Anim Genet.* 2015; 46: 466. doi:10.1111/age.12313

Declaration of Originality

Last name, first name: Henkel, Jan

Matriculation number: 17-133-521

I hereby declare that this thesis represents my original work and that I have used no other sources except as noted by citations.

All data, tables, figures and text citations which have been reproduced from any other source, including the internet, have been explicitly acknowledged as such.

I am aware that in case of non-compliance, the Senate is entitled to withdraw the doctorate degree awarded to me on the basis of the present thesis, in accordance with the "Statut der Universität Bern (Universitätsstatut; UniSt)", Art. 69, of 7 June 2011.

Place, date

Bern, 02.03.2021

Signature

Jan Henkel Digitally signed by Jan Henkel
Date: 2021.03.02 15:33:29 +01'00'